PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/00, C12N 1/15, 1/21, 5/10, 15/12, 15/63 — A1

(11) International Publication Number: **WO 99/35158**

(43) International Publication Date: 15 July 1999 (15.07.99)

(21) International Application Number: PCT/US99/00108

(22) International Filing Date: 6 January 1999 (06.01.99)

(30) Priority Data:

60/070,657	7 January 1998 (07.01.98)	US
60/070,658	7 January 1998 (07.01.98)	US
60/070,692	7 January 1998 (07.01.98)	US
60/070,704	7 January 1998 (07.01.98)	US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 3142 Quesada Street, N.W., Washington, DC 20015 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). BREWER, Laurie, A. [US/US]; Apartment 115, 410 Van Dyke Street, St. Paul, MN 55119-4321 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). DUAN, Roxanne, D. [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US).

(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: 36 HUMAN SECRETED PROTEINS

(57) Abstract

The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

36 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein , a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched , for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene was shown to have homology to the human T-cell receptor interacting molecule (TRIM) protein which is thought to be important in modulating the immune response to antigens. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: IKISLKKRS (SEQ ID NO:101). Polynucleotides encoding these polypeptides are also encompassed by the invention. When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene

activates promyeloid cells, or more generally, immune or other cells or cell types through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in normal, apoptotic, and transformed T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, inflammatory, and neoplastic conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, inflammatory, neoplastic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:56 as residues: Glu-35 to Asp-53, Met-82 to Gln-107, Val-117 to Gly-125.

The tissue distribution in apoptotic and transformed T-cells, combined with the homology with the TRIM protein and detected GAS biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of inflammation, blood disorders and neoplasms. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions, particularly in T-cell directed immune responses. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne,

neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1699 of SEQ ID NO:11, b is an integer of 15 to 1713, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in placenta, and to a lesser extent, in ovarian and testis tumor, and in T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, developmental, reproductive, vascular and/or neoplastic conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, developmental, reproductive, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in testis and ovarian tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of immune and reproductive disorders and cancer. Moreover, since the gene is also expressed in T-cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues.

In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the protein is useful in the detection, treatment, and/or prevention of vascular conditions, which include, but are not limited to, microvascular disease, vascular leak syndrome, aneurysm, stroke, atherosclerosis, arteriosclerosis, or embolism. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2048 of SEQ ID NO:12, b is an integer of 15 to 2062, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene was found to have homology to the conserved Homo sapiens mRNA for a human 36 kDa phosphothyrosine protein (See Genbank Accession No.gb|AJ223280|HSAJ3280, and J. Exp Med 1998 Apr 6;187(7):1157-61 which is hereby incorporated by reference) which is thought to be involved in modulation of the immune response, potentially in signaling events during T and NK cell activation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GTRGLSTVSWTHTQPSKRGDPSREPRGGHSCLLPGSPATWCLPAPCSLPGPVL TPSSSGLDSALEGPRGA ASLLRAPLQ (SEQ ID NO:102), HTQPSKRGDPSREP RGGHSCLLP (SEQ ID NO:103), and/or VLTPSSSGLDSALEGPRGA ASL (SEQ ID NO:104). Polynucleotides encoding these polypeptides are also encompassed by the invention. When tested against Jurket cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates T-cells, or more generally immune cells or other cells and cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in various T-cell and macrophage lines, and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune or hematopoietic disorders, particularly inflammation, infections, or immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:58 as residues: Pro-33 to His-49, Glu-74 to Lys-83.

The tissue distribution in T-cells and macrophages, combined with the homology to the conserved phosphothyrosine protein and detected GAS biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of inflammatory and general immune disorders. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1210 of SEQ ID NO:13, b is an integer of 15

to 1224, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: AGSRTNNEQIE (SEQ ID NO:105). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in aortic endothelial, cardiac, and adipose cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic, and vascular conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., metabolic, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in aortic endothelial and cardiac tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cardiovascular and metabolic disorders. More generally, the protein is useful in the detection, treatment, and/or prevention of vascular conditions, which include, but are not limited to, microvascular disease, vascular leak syndrome, aneurysm, stroke, atherosclerosis, arteriosclerosis, or embolism. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1607 of SEQ ID NO:14, b is an integer of 15

to 1621, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

The translation product of this gene was found to have homology to the human antigen NY-CO-38 (See Genbank Accession No.gil3170200 (AF039700)) which is thought to be important in modulation of the immune response. Polynucleotides of the invention do not comprise the polynucleotide sequence shown as Genbank Accession No.gil3170200, which is hereby incorporated herein by reference. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GTSTSSRGRLHACGHS (SEQ ID NO:106), PSSEVQKGKPNSPLGNS ELRPHLVNTKPRTSLERGHTIPFLWPSEFGLSQLWGTPSLNPNKTPLESLSLH PSPLPSALIAARIVTPNLTLSSLIK (SEQ ID NO:107), PNSPLGNSELRPHLV NTKPRT (SEQ ID NO:108), LSLHPSPLPSALIAARIVTPNLT (SEQ ID NO:109), PGSQGAAAGRELFMTDRERLAEARQRELQRQELLMQKRLAMESNKILQEQ QEMERQRRKEIAQKAAEENERYRKEMEQIVEEEEKFKKQWEEDWGSKEQLLLP KTITAEVHPVPLRKPKYDQGVEPELEPADDLDGGTEEQGEQDFRKYEEGFD PYSMFTPEQIMGKDVRLLRIKKEGSLDLALEGGVDSPIGKVVVSAVYERGAA ERHGGIVKGDEIMAINGKIVTDYTLAEADAALQKAWNQGGDWIDLVVAVC PPKEYDDELTFF (SEQ ID NO:110), GRELFMTDRERLAEARQRELQRQ (SEQ ID NO:111), QQEMERQRRKEIAQKAAEENER (SEQ ID NO:112), KPKYDQGVEP ELEPADDLDGGTEEQ (SEQ ID NO:113), and/or IVTDYTLAEADAALQKAWN QGGDWI (SEQ ID NO:114). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in epididiymus, and, to a lesser extent, in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive or immune disorders, particularly autoimmune conditions, and/or infertility. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell

types (e.g., reproductive, immune, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:60 as residues: Pro-18 to Lys-26.

The tissue distribution in the epididiymus and monocytes, combined with the homology to the NY-CO-38 antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalties of the epididiymus, including infertility. Moreover, the protein may be used in the development of contraceptives, modulating the immune response (e.g. activating or inhibiting), or inflammatory conditions. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1406 of SEQ ID NO:15, b is an integer of 15 to 1420, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune or hematopoietic disorders, particularly acute inflammatory reactions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of

the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:61 as residues: Gly-18 to Ser-27, Gly-46 to Asp-51.

The tissue distribution in neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of neutrophil mediated disease such as septic shock and acute inflammatory conditions. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1021 of SEQ ID NO:16, b is an integer of 15 to 1035, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

This gene is expressed primarily in brain frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neural disorders, particularly diseases of the central nervous system (CNS) and brain frontal cortex. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:62 as residues: Asn-17 to Arg-29, Gly-37 to Cys-45.

The tissue distribution in brain frontal cortex indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, diagnosis and treatment of diseases of the CNS. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is

involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 844 of SEQ ID NO:17, b is an integer of 15 to 858, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:17, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HTML PLKIAAPYLLENCSCPIYISTSPHLFLST (SEQ ID NO:115). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 20. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 20.

This gene is expressed primarily in T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune or hematopoietic disorders, particularly leukemias, lymphomas, hematopoeitic disorders, auto-immunities and immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a

disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:63 as residues: His-31 to Ala-40.

The tissue distribution in T cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-supressive conditions (transplantation) and hematopoeitic disorders. In addition this gene product may be applicable in conditions of general microbial infection, inflammation or cancer. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 922 of SEQ ID NO:18, b is an integer of 15 to

936, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: FSILFAFVLFYPGSFFTLP (SEQ ID NO:116). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pineal gland, bone marrow, fetal liver and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine, immune, hematopoietic, or reproductive disorders, particularly cancers of the blood forming cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endocrine, immune, hematopoietic, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:64 as residues: Cys-29 to Ser-41.

The tissue distribution in bone marrow and fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for disorders of the blood including lymphomas and leukemias. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as

infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the protein is useful in the detection, treatment, and/or prevention of a variety of endocrine or reproductive disorders, particularly infertility, and biological clock and ion homeostasis disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 599 of SEQ ID NO:19, b is an integer of 15 to 613, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

The translation product of this gene shares sequence homology with the kidney injury molecule-1 of Rattus norvegicus which is thought to play an important role in the restoration of the morphological integrity and function to postischemic kidney (See Genbank Accession No. gil2665892 (AF035963)). Polynucleotides and polypeptides of the present invention are useful to promote growth of new tissue and survival of damaged tissue. Recombinant polypeptides of the present invention can be expressed in prokaryotic and eukaryotic host cells using a claimed process. Soluble variants fused to a toxin, imageable compound or radionuclide, and IgG fusion proteins are also claimed.

Polynucleotides and polypeptides of the present invention or an agonist, can be used to treat renal disease and to promote the growth of new tissue or the survival of damaged tissue, generally in conditions where the binding of specific ligand to the present invention stimulates cell growth, maintains cellular differentiation or reduces apoptosis, e.g. in cases of renal failure, nephritis, kidney transplants, toxic or hypoxic injury. A monoclonal antibody specific for polynucleotides and polypeptides of the present invention can be used to treat renal disease, e.g. where binding of the invention

to a ligand results in neoplasia, loss of cellular function, susceptibility to apoptosis or promotion of inflammation, deliver imaging agents to the cells expressing the present invention in vivo or in vitro and measure the concentration of the present invention by immunoassay. Damage/regeneration of renal cells can be determined by measuring the present invention, particularly to diagnose or monitor the progress of disease or therapy.

The tumour cells expressing the present invention can be inhibited by treatment with a fusion protein comprising a ligand of the present invention or MAb with a toxin or radionuclide, and tumour cells that express the present invention ligand can be inhibited with similarly tagged polypeptides of the present invention or anti-present invention ligand antibody. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HESTVK (SEQ ID NO:117). Polynucleotides encoding these polypeptides are also encompassed by the invention.

It has been discovered that this gene is expressed primarily in infant brain and fetal liver, and to a lesser extent in neoplastic cell lines and endocrine organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, neural, endorcine, and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, neural, endorcine, growth, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:65 as residues: Ser-44 to Ser-51, Cys-53 to Cys-64, Val-76 to Lys-83, Pro-102 to Gly-108, Arg-133 to Thr-162, Thr-169 to Lys-183.

The tissue distribution in fetal liver and infant brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of immune and developmental conditions. Moreover, the expression within infant and fetal tissues and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility

in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Alternatively, the protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritus, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 557 of SEQ ID NO:20, b is an integer of 15 to 571, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LENLGTHKKKDSFSVKTVGICCCFHLN (SEQ ID NO:118). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 14. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 14.

This gene is expressed primarily in infant brain, and to a lesser extent, in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neural, immune, or developmental disorders, particularly abnormalities of the infant brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, immune, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain and monocytes indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the central nervous system. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein may also be useful in modulating the immune response towards self-antigens.

In addition, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2010 of SEQ ID NO:21, b is an integer of 15 to 2024, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: FTKCFH (SEQ ID NO:119). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune or hematopoietic disorders, particularly mononucleosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:67 as residues: Ser-27 to Arg-39, Pro-91 to Arg-100.

The tissue distribution in monocytes indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of monocytes such as mononucleosis. Moreover, the protein product of this

gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:22 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 561 of SEQ ID NO:22, b is an integer of 15 to 575, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:22, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: QNMNDYNI (SEQ ID NO:120). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in a bone marrow cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune or hematopoietic disorders, particularly immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be

routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:68 as residues: Arg-26 to Trp-32.

The tissue distribution in bone marrow cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities associated with the bone marrow, such as immunodeficiencies. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1167 of SEQ ID NO:23, b is an integer of 15 to 1181, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:
PARHLWTPSPVCKPSIKPHVSFAGSGSLWRLEPYAFPIEVNRGSAQHWVPG (SEQ ID NO:121), and/or VCKPSIKPHVSFAGSGSLWRLEPYAFPIE (SEQ ID NO:122). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neural, metabolic, or developmental disorders, particularly fetal immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, metabolic, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:69 as residues: Met-24 to Arg-29.

The tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of abnormalities of the fetal liver such as fetal immunodeficiencies. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia, since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

Alternatively,the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1751 of SEQ ID NO:24, b is an integer of 15 to 1765, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: QFSFLSAKGLHWALFVFFYFLSTACQRWAWGL (SEQ ID NO:123). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in placenta, breast, testes, fetal liver, and to a lesser extent, in endocrine organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and hormonal conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gonadal and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, seminal fluid, breast milk, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:70 as residues: Thr-15 to Trp-21, Val-33 to Gln-39.

The tissue distribution placenta, breast, and testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of reproductive and hormonal disorders. Moreover, the protein is useful in the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sachs disease, phenylkenonuria, galactosemia, hyperlipidemias, porphyrias, and Hurler's syndrome. In addition, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 877 of SEQ ID NO:25, b is an integer of 15 to

891, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

When tested against Jurket cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates T-cells, or more generally, immune cells or other cells or cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in CD34-depleted cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and hemopoietic conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hemopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:71 as residues: Gln-26 to His-34.

The tissue distribution in CD34-depleted cord blood, combined with the detected GAS biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of immunological and blood disorders. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia,

leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 451 of SEQ ID NO:26, b is an integer of 15 to 465, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The gene encoding the disclosed cDNA is believed to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

This gene is expressed primarily in prostate cancer, and to a lesser extent, in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive or neural disorders, particularly prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues or cell types (e.g., reproductive, neural, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, seminal fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:72 as residues: His-34 to Asn-40.

The tissue distribution in prostate and brain tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of prostate cancer. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. In addition, the protein is useful in the development of novel contraceptives. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 769 of SEQ ID NO:27, b is an integer of 15 to 783, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HEPGRCGPENLALQATQRGTRFSVPMCKSSRQYTYTSVHMCQCACERVEWRG SLTPARALHNHLTEQWFP HGFPFLSRFFTY (SEQ ID NO:124), ENLALQATQRG TRFSVPMCKSSRQ (SEQ ID NO:125), and/or MCQCACERVEWRGSLTPARALH NHLT (SEQ ID NO:126). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal brain and CD34 depleted cord blood. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neural, immune, or reproductive conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, immune, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:73 as residues: Ser-35 to Asp-48.

The tissue distribution in fetal brain and CD34 depleted cord blood indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of nervous system and immune disorders. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral

neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein is also useful in the modulation of the immune response, particularly auto-antigens.

The expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 456 of SEQ ID NO:28, b is an integer of 15 to 470, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LRRASCPIWSKD (SEQ ID NO:127). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on the X chromosome. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for the X chromosome.

This gene is expressed primarily in fetal liver spleen and retina, and to a lesser extent in, placenta and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, hemopoietic, retinal, or vascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the retinal, hemopoietic and placental tissues and systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., developmental, hemopoietic, retinal, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, bile, pulmonary surfactant and sputum, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution fetal liver spleen and retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, diagnosis and treatment of retinal, hemopoietic, and developmental disorders. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The protein is also useful in the modulation of the immune response, particularly to auto-antigens in autoimmune disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the

scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1307 of SEQ ID NO:29, b is an integer of 15 to 1321, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene shares sequence homology with the flatfish growth hormone (see Genbank gb|E02221|E02221) which is thought to be important in growth regulation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GTSTKLPYCRENVCLAYGSEWSVYAVGSQAHVSFLDPRQPSYNVKSVCSRER GSGIRSVSFYEHIITVGTGQGSLLFYDIRAQRFLEERLSACYGSKPRLAGENLK LTTGKGWLNHDETWRNYFSDIDFFPNAVYTHCYDSSGTKLFVAGGPLPSG LHGNYAGLWS (SEQ ID NO:128), CRENVCLAYGSEWSVYAVGSQA (SEQ ID NO:129), PSYNVKSVCSRERGSGIRSVSFYE (SEQ ID NO:130), DIRAQRFLEER LSACYGSKPRLAGENLKL (SEQ ID NO:131), KLTTGKGWLNHDETWRNYF SDIDFFP (SEQ ID NO:132), and/or YDSSGTKLFVAGGPLPSGLHG (SEQ ID NO:133). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in infant brain, fetal liver spleen and to a lesser extent, in pancreas tumor and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth, developmental, immune, neural, or metabolic disorders, particularly cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and hemopoietic tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., growth, developmental, neural, immune, metabolic, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken

from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:75 as residues: Gly-31 to Pro-37, His-46 to Ala-52.

The tissue distribution in infant brain, combined with the homology to flatfish growth hormone indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, treatment and diagnosis of growth disorders such as cancer particularly pancreas tumor. Moreover, the distribution in fetal liver spleen and bone marrow indicates that the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

In addition, the expression within fetal and infant tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:30 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 606 of SEQ ID NO:30, b is an integer of 15 to

620, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:30, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

The translation product of this gene shares sequence homology with Dictyostelium discoideum random slug cDNA 25 protein. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: KPQRFRRPFFFNHPKPSSHPGLHSRPTLHSHPAFHSHPELQQPTQTSPVPLTPE SPLFQNFSGYHIGVGRADCTGQVADINLMGYGKSGQNAQGILTRLYSRAFI MAEPDGSNRTVFVSIDIGMVSQRLRLEVLNRLQSKYGSLYRRDNVILSGTHT HSGPAGYFQYTVFVIASEGFSNQTFQHMVTGILKSIDIAHTNMKPGKIFINK GNVDGVQINRSPYSYLQNPQSERARYSSNTDKEMIVLKMVDLNGDDLGLISFS FSKSALGTYYEPRNTSLE (SEQ ID NO:134), KPSSHPGLHSRPTLHSHPAFHS HPELQQPT (SEQ ID NO:135), RADCTGQVADINLMGYGKSGQNAQGI (SEQ ID NO:136), RAFIMAEPDGSNRTVFVSIDIGMV (SEQ ID NO:137), RLQSKYGSLYR RDNVILSGTHTHSGPA (SEQ ID NO:138), ASEGFSNQTFQHMVTGILKSIDI (SEQ ID NO:139), IFINKGNVDGVQINRSPYSYLQNP (SEQ ID NO:140), and/or TDKEMIVLKMVDLNGDDLGLISFSFSKSAL (SEQ ID NO:141). Polynucleotides encoding these polypeptides are also encompassed by the invention. When tested against Jurket cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates T-cells, or more generally, immune cells or other cells or cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in liver (hepatoma).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hepatic disorders, particularly liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:76 as residues: Glu-39 to Gly-45, Thr-51 to Gly-60, Ala-63 to Gln-82.

The tissue distribution in liver, combined with the homology to the Dictyostelium discoideum random slug cDNA 25 protein and the detected GAS biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, treatment and diagnoisis of liver diseases and disorders. Moreover, the protein product of this gene is useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:31 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1180 of SEQ ID NO:31, b is an integer of 15 to 1194, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:31, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HLTQFCVIQLLPTHL (SEQ ID NO:142), MLASXSVLCDPAP ANPSDELLVHSPCFLLSPPVAPVPFFPCAKLIPAPRPVRYFSPPDLRLGNTPAP NSEITYTPSVHFCHPIAKLLCLKVRNLCEGVLSAAFPKA (SEQ ID NO:143), ANPSDELLVHSPCFLLSPPVAPVPFFP (SEQ ID NO:144), and/or FSPPDLRLGNT

PAPSEITYTPSVHFCHPIAKLLC (SEQ ID NO:145). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in liver (hepatoma), fetal dura mater and fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver and central nervous system (CNS) diseases or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system and CNS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., hepatic, neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, bile, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:77 as residues: Leu-18 to His-23, Pro-25 to Asp-32.

The tissue distribution in liver, fetal dura mater, and fetal brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, diagnosis and treatment of diseases of the liver and CNS. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival.

The protein is also useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:32 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 815 of SEQ ID NO:32, b is an integer of 15 to 829, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:32, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LIFHFVYL (SEQ ID NO:146), PTPRVILQVGSRIADRVYDIPR NFPLALDLGCGRGYIAQYLNKLQLFHCRKLLESFSKLTLQKMLCLHWVNDL PRALEQIHYILKPDGVFIGAMFGGDTLYELRCSLQLAETEREGGFSPHISPFTAV NDLGH LLGRAGFNTLTVDTDEIQVNYPGMFELMEDLQEQKSRMLT (SEQ ID NO:147), LQVGSRIADRVYDIPRNFPLALDL (SEQ ID NO:148), GYIAQYLNK LQLFHCRKLLESFSK (SEQ ID NO:149), VNDLPRALEQIHYILKPDGVFIGAMFG (SEQ ID NO:150), YELRCSLQLAETEREGGFSPHISPFTAV (SEQ ID NO:151), and/or NTLTVDTDEIQVNYPGMFELME (SEQ ID NO:152). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 20. Accordingly,

polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 20.

This gene is expressed primarily in eosinophils and fetal lung, and to a lesser extent, in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune, hemopoietic and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hemopoietic and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hemopoietic, respiratory, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, pulmonary surfactant and sputum, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in eosinophils and bone marrow indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, hemopoietic and respiratory diseases. Moreover, the protein product of this gene is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:33 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the

scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1322 of SEQ ID NO:33, b is an integer of 15 to 1336, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:33, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with C. elegans transient receptor potential protein and with human O3O gene (PCT patent application, publication no. WO9630389), which is thought to be differentially expressed on tumor cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:
MRQLFYDPDECGLMKKGGLYFSDFWNKLDVGAILLFVAGLTCRLIPATLYPGR VILSLDFILFCLRLMHI FTISKTLGPKIIIVKR (SEQ ID NO:153), DECGLMKKGG LYFSDFWNKLDVGAILL (SEQ ID NO:154), TLYPGRVILSLDFILFCLRLMHIFT (SEQ ID NO:155), VPRERRDAAEPAFPEWLTVLLLCLYLLFTNILLLNLLIAM FNYTFQQVQEHTDQIWKFQRHDLIEEYHGRPAVPPPLILFSHLQLFIKRVVL KTPAKRHKQLKNKLEKNEEAALLSWEIYLKENYLQNRQFQQKQRPEQKIEDI SNKVDAMVDLLDLDPLKRSGSMEQRLASLEEQVAQTARALHWIVRTLRASG FSSEADVPTLASQKAAEEPDAEPGGRKKTEEPGDSYHVNARHLLYPNCPVTRF PVPNEKVPWETEFLIYDPPFYTAERK (SEQ ID NO:156), QIWKFQRHDLIEEYH GRPAVPPPLILFS (SEQ ID NO:157), LQNRQFQQKQRPEQKIEDISNKVDAMVD (SEQ ID NO:158), VPTLASQKAAEEPDAEPGGRKKTE (SEQ ID NO:159), and/or PNEKVPWETEFLIYDPPFYT (SEQ ID NO:160). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in dendritic cells and adult brain, and to a lesser extent, in cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neural, immune, or hematopoietic disorders, particularly of the central nervous system (CNS) and diseases of the brain, such as tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the CNS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:79 as residues: Pro-41 to Ala-55.

The tissue distribution in adult brain and cerebellum, combined with the homology to C. elegans transient receptor potential protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of disorders of the CNS such as tumors. Moreover, the protein product of this gene is useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival.

Alternatively, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus

erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:34 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1300 of SEQ ID NO:34, b is an integer of 15 to 1314, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:34, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PESFNFCFGPGVPMPWCLLPVLSVLHWSTEDTRSCGAQGGGPPLPPRG (SEQ ID NO:161). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is thought to reside on chromosome 11. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 11.

This gene is expressed primarily in infant brain and osteoblastic tissue, and to a lesser extent in leukocytic cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurodegenerative diseases of the brain, osteosarcoma, osteoporosis and osteonecrosis, hematopoeitic disorders and other immune deficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,

particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, neurodegenerative, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, immune, and osteoblast tissues indicates that the protein product of this gene is useful for the treatment and diagnosis of conditions affecting the central nervous, skeletal and immune systems. Expression in the infant brain indicates a potential role in developmental and neurodegenerative diseases of the brain, as well as behavioral or other nervous system disorders, such as depression, schizophrenia, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders. Expression of this gene product in osteoblastic tissue would suggest a role in the treatment and diagnosis of osteoperosis, fracture, osteosarcoma, ossification, osteonecrosis, arthritis and trauma. Immune and hematopoietic disorders of potential applicability would include leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS) and immuno-supressive conditions (transplantation). In addition this gene product may be applicable in conditions of general microbial infection, inflammation and cancer. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:35 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1250 of SEQ ID NO:35, b is an integer of 15 to 1264, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:35, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

This gene is expressed primarily in endometrial tissue and to a lesser extent in breast cancer tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the following diseases and conditions, which include, but are not limited to, developmental anomalies, fetal deficiencies, female infertility, ovarian, breast and endometrial cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endometrium, breast, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in endometrium and breast cancer tissues indicates that the protein product of this gene is useful for the treatment and diagnosis of developmental anomalies or fetal deficiencies, as well as breast, ovarian and endometrial cancers, in addition to cancers of other tissues where expression is observed. Furthermore, the tissue distribution indicates that the protein product of this gene is useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelyhood of implantation and development of a healthy embryo. In both cases this gene or its gene product could be administered at later stages of pregnancy to promote heathy development of the endometrium. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:36 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 674 of SEQ ID NO:36, b is an integer of 15 to 688, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:36, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

The translation product of this gene shares sequence homology with amyloid precursor protein protease which is thought to be important in the processing or clearance of amyloid precursor protein to form beta-amyloid peptide. The translation product of this gene also shares significant homology with Enamel Matrix Serine Proteinase 1 from Sus scrofa.

This gene is expressed primarily in keratinocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, skin cancers (melanomas), eczema, psoriasis or other disorders of the skin, as well as amyloid protein related conditions such as transmissible spongiform encephalopathy (TSE), Creutzfeldt-Jakob disease (CJD) and Alzheimer's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin and brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., skin, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:82 as residues: Asn-31 to Thr-41, Pro-43 to Asp-49, Glu-56 to Arg-66, Ser-71 to Trp-80, Asn-160 to Val-169, Thr-192 to Val-198, Lys-215 to Asp-226, Asp-234 to Gly-246, Pro-265 to Gly-273.

The tissue distribution in skin indicates that the protein product of this gene is useful for the treatment and diagnosis of skin cancers (melanomas), eczema, psoriasis or other disorders of the skin. The homology to amyloid precursor protein-cleaving protease indicates that this gene product cold have a role in the processing or clearance

of amyloid precursor protein to form beta-amyloid peptide and therefore useful for treating or preventing conditions associated with beta-amyloid peptide such as Alzheimer's disease, transmissible spongiform encephalopathy (TSE) and Creutzfeldt-Jakob disease (CJD). Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:37 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1502 of SEQ ID NO:37, b is an integer of 15 to 1516, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:37, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LWTMNPASDGGTSESIFDLDYASWGIRSTL (SEQ ID NO:162). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pulmonary and endothelial tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, asthma, pulmonary edema, atherosclerosis, restenosis, stroke potential, thrombosis and hypertension, Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and cardiovascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., vascular tissues, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in pulmonary and endothelial tissues indicates that the protein product of this gene is useful for the treatment and diagnosis of cardiovasular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis hypertension, inflammation and wound healing. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:38 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1253 of SEQ ID NO:38, b is an integer of 15 to 1267, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:38, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

The translation product of this gene shares sequence homology with cathepsin b, a cysteine protease which is thought to be important in demyelination, emphysema, rheumatoid arthritis, and neoplastic infiltration. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GTSGGARASLGPSPHLHQGPGAT (SEQ ID NO:163). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, demyelination, emphysema, rheumatoid arthritis, neoplastic infiltration, atherosclerosis, restenosis, thrombosis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endothelium, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endothelial tissues, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:84 as residues: Gln-23 to Glu-30, Ala-43 to Leu-53, Tyr-118 to Gln-137, Gln-152 to Trp-157.

The homology to the cysteine protease cathepsin b indicates that the protein product of this gene is useful for the treatment and diagnosis of pathologies such as demyelination, emphysema, rheumatoid arthritis and neoplastic infiltration. In addition, the expression of this gene in endothelial tissues and cells indicates that it may be useful in the diagnosis and treatment of cardiovascular and pulmonary conditions such as asthma, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis, hypertension, inflammation and wound healing. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:39 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2189 of SEQ ID NO:39, b is an integer of 15 to 2203, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:39, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 30

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells, or more generally, immune cells, in addition to other cells or cell types, through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the

Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: YLVIFFLKC (SEQ ID NO:164). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta, spleen and a variety of blood cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and hematopoietic disorders, as well as those affecting reproduction and fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are useful to provide immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in immune tissues indicates that the protein product of this gene is useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-supressive conditions (transplantation) and hematopoeitic disorders. In addition, this gene product may be applicable in conditions of general microbial infection, inflammation or cancer. Furthermore, the expression in placenta would suggest a potential application in the treatment and diagnosis of developmental anomalies or fetal deficiencies. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus.

Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells

or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:40 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1712 of SEQ ID NO:40, b is an integer of 15 to 1726, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:40, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GSQVNGV (SEQ ID NO:165). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, auto-immunities, immunodeficiencies, hematopoeitic disorders, general microbial infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:86 as residues: Gly-15 to Gly-21, Pro-32 to Trp-38.

The tissue distribution in neutrophils indicates that the protein product of this gene is useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-supressive conditions (e.g. transplantation) and hematopoeitic disorders. In addition this gene product may be applicable in conditions of general microbial infection, inflammation and cancer. Furthermore, expression of this gene product in neutrophils also strongly indicates a role for this protein in immune function and immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:41 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1088 of SEQ ID NO:41, b is an integer of 15 to 1102, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:41, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 32

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: RPTRPLNCGR (SEQ ID NO:166). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is thought to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed in a wide variety of tissues suggesting its presence in the vasculature. Those tissues include tumors such as rhabdomyosarcoma, hepatoma, hepatocellular carcinoma uterine cancer, hemangiopericytoma, lymphoma, and testes tumor, but it is also present in T-cell fractions, dendritic cells, endothelial cells, bone marrow stromal cells and melanocytes melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, treatment of highly vascularized tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., vascular tissues, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:87 as residues: Pro-56 to Pro-63, Met-92 to Thr-98, Ser-112 to Pro-120, Pro-162 to Ser-169.

The tissue distribution in vascular tissues indicates that the protein product of this gene is useful for treating disorders of the vasculature and highly vascularized tumors, as well as tumors of other tissues where expression has been observed. Furthermore, the tissue distribution in vascular tissues indicates that the protein product of this gene is useful for the diagnosis and treatment of conditions and pathologies of the cardiovascular system, such as heart disease, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:42 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1383 of SEQ ID NO:42, b is an integer of 15 to 1397, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:42, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

The gene encoding the disclosed cDNA is thought to reside on chromosome 11. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 11.

This gene is widely expressed including expression in several libraries from fetal tissues. The gene is also expressed in placenta, lung, synovium, bone marrow, testes and several regions of the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vascular degenerative conditions or tumor associated neo-vascularization. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., vascular, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:88 as residues: Asp-40 to Glu-50, Ser-59 to Gly-69, Ala-98 to His-105, Arg-108 to Glu-114, Pro-124 to Ser-138, Ala-143 to Gly-154.

The tissue distribution in vascular tissues indicates that the protein product of this gene is useful for treating conditions pertaining to abnormal vascular states such as certain retinopathies, diabetes induced eschemia, or vascularization associated with certain types of tumors. Furthermore, the translation product of this gene is useful in the treatment and diagnosis of conditions and pathologies of the cardiovascular system, such as heart disease, restenosis, atherosclerosis, stoke, angina, thrombosis, and wound healing. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more

generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:43 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1725 of SEQ ID NO:43, b is an integer of 15 to 1739, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:43, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

The translation product of this gene shares homology with a C. elegans protein. (>sp|Q11073|YT45). In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:
VWGPPSVAAALELVDPPGCREFGTSNIEDRDELAYHISI (SEQ ID NO:167).
Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is thought to reside on chromosome 14. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 14. When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells, or more generally, immune cells, or other cells or cell types, through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in liver and fetal liver, and to a lesser extent in several other tissue and organs including tissue from Alzheimers disease.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic, liver and neural diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and neural systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., liver, spleen, neurological, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:89 as residues: Lys-21 to Gln-32, Asp-117 to Glu-124, Tyr-179 to Gly-184, Asn-211 to Gly-217, Leu-239 to Lys-264.

The tissue distribution in fetal liver/spleen and Alzheimer's tissues indicates that the protein product of this gene is useful for the diagnosis and treatment of metabolic and neural diseases and cancer. Expression of this gene product in fetal liver/spleen indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also,used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:44 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3047 of SEQ ID NO:44, b is an integer of 15 to 3061, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:44, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in brain medulloblastoma, and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain medulloblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., brain, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:90 as residues: Pro-53 to Thr-65.

The tissue distribution in brain medulloblastoma indicates that the protein product of this gene is useful for the diagnosis and treatment of brain medulloblastoma and other disorders of the CNS, as well as cancers of other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:45 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 960 of SEQ ID NO:45, b is an integer of 15 to 974, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:45, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

The translation product of this gene shares sequence homology with mouse oligodendrocyte specific protein and Clostridium endotoxin receptor which is thought to be important in host response to bacterial infections. The translation product of this gene also shares significant homology with the conserved claudin-1 protein of Mus musculus, which is thought to be located with tight junctions, but the function of the mouse gene has not yet been deduced. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LPPRGPATFGSPGCPPANSPPSAPATPEPARAPERV (SEQ ID NO:168). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thymic stromal cells, heart, dendritic and colon carcinoma cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious conditions and neoplasms. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:91 as residues: Ala-54 to Ala-64.

The tissue distribution in immune and gastrointestinal tissues indicates that the protein product of this gene is useful for the study and treatment of immune, infectious, and cancerous disorders. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:46 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1432 of SEQ ID NO:46, b is an integer of 15 to 1446, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:46, and where b is greater than or equal to a + 14.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HTGAU75	209563 12/18/97	Uni-ZAP XR	11	1713	1	1713	149	149	56	1	33	34	142
2	HTTDP47	209563 12/18/97	Uni-ZAP XR	12	2062	1	2062	133	133	57	1	21	22	50
3	HTXJQ11	209563 12/18/97	Uni-ZAP XR	13	1224	1	1224	238	238	58	1	29	30	101
4	HADCO45	209563 12/18/97	pSport1	14	1621	1	1621	187	187	59	1	19	20	47
5	HMIAL37	209563 12/18/97	Uni-ZAP XR	15	1420	1	1420	49	49	60	1	13	14	97
6	HNGDU40	209563 12/18/97	Uni-ZAP XR	16	1035	1	1035	333	333	61	1	17	18	51
7	HFXBO84	209563 12/18/97	Lambda ZAP II	17	858	1	858	9	9	62	1	18	19	51

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
8	HLLAX19	209563 12/18/97	pCMVSport 1	18	936	1	936	298	298	63	1	22	23	40
9	HPMAG94	209563 12/18/97	Uni-ZAP XR	19	613	1	613	160	160	64	1	31	32	41
10	HSVAK93	209563 12/18/97	Uni-ZAP XR	20	571	1	571	21	21	65	1	24	25	183
11	HMQBO88	209563 12/18/97	Uni-ZAP XR	21	2024	1	2024	141	141	66	1	31	32	57
12	HMQBU45	209563 12/18/97	Uni-ZAP XR	22	575	1	575	275	275	67	1	24	25	100
13	HMWAJ53	209563 12/18/97	Uni-Zap XR	23	1181	1	1181	238	238	68	1	17	18	73
14	HNGBQ56	209563 12/18/97	Uni-ZAP XR	24	1765	1	1765	155	155	69	1	21	22	48
15	HAUAU73	209563 12/18/97	Uni-ZAP XR	25	891	1	869	316	316	70	1	25	26	45

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
16	HCUGO12	209563 12/18/97	ZAP Express	26	465	1	465	133	133	71	1	25	26	63
17	HPFCX44	209563 12/18/97	Uni-ZAP XR	27	783	1	783	85	85	72	1	23	24	47
18	HCUBV79	209563 12/18/97	ZAP Express	28	470	1	470	203	203	73	1	21	22	48
19	HLQBV04	209563 12/18/97	Lambda ZAP II	29	1321	1	1321		268	74	1	23	24	46
20	HMADW66	209563 12/18/97	Uni-ZAP XR	30	620	1	620	60	60	75	1	19	20	58
21	HLDBE54	209563 12/18/97	pCMVSport 3.0	31	1194	1	1194	130	130	76	1	26	27	89
22	HFTAB66	209563 12/18/97	Uni-ZAP XR	32	829	21	762	72	72	77	1	21	22	43
23	HEOMQ63	209563 12/18/97	pSport1	33	1336	1	1336	123	123	78	1	23	24	47

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
24	HDPJM30	209563 12/18/97	pCMVSport 3.0	34	1314	1	1313	259	259	79	1	20	21	59
25	HJABT47	209563 12/18/97	pBluescript SK-	35	1264	231	1264	324	324	80	1	36	37	47
26	HJMAG88	209563 12/18/97	pCMVSport 3.0	36	688	1	688	20	20	81	1	23	24	47
27	HKAAH36	209563 12/18/97	pCMVSport 2.0	37	1516	1	1516	254	254	82	1	29	30	293
27	HKAAH36	209563 12/18/97	pCMVSport 2.0	51	1381	196	1381	129	129	96	1	29	30	293
27	HKAAH36	209563 12/18/97	pCMVSport 2.0	52	1439	1	1439	189	189	97	1	29	30	61
28	HMADS41	209563 12/18/97	Uni-ZAP XR	38	1267	1	1267	267	267	83	1	21	22	88
29	HMEFT85	209563 12/18/97	Lambda ZAP II	39	2203	1	2203	70	70	84	1	23	24	249

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
30	HMSBX80	209563 12/18/97	Uni-ZAP XR	40	1726	1	1726	169	169	85	1	19	20	57
31	HNGCL23	209563 12/18/97	Uni-ZAP XR	41	1102	1	1102	269	269	86	1	18	19	48
32	HNTBI26	209563 12/18/97	pCMVSport 3.0	42	1397	1	1397	32	32	87	1	35	36	172
32	HNTBI26	209563 12/18/97	pCMVSport 3.0	54	1368	1	1368	16	16	99	1	35	36	131
33	HPIBO15	209563 12/18/97	Uni-ZAP XR	43	1739	1	1739	127	127	88	1	18	19	173
34	HCYBG92	209563 12/18/97	pBluescript SK-	44	3061	1	2661	118	118	89	1	21	22	274
35	HMDAQ29	209563 12/18/97	Uni-ZAP XR	45	974	1	974	180	180	90	1	48	49	82
36	HSYBI49	209563 12/18/97	pCMVSport 3.0	46	1446	1	1446	230	230	91	1	21	22	70

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification , such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown inTable 1, the ORF (open reading frame), or any fragement specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identiy are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not, matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-

60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the

polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat

polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying

personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors),or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, caricadic rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSport 2.0	pCMVSport 2.0
pCMVSport 3.0	pCMVSport 3.0
pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P^{32} using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D.^{600}$) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or , Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 μm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem

columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring

signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μg of a plasmid containing the polynucleotide is co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One μg of BaculoGold™ virus DNA and 5 μg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture

and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ^{35}S-methionine and 5 µCi ^{35}S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No.209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:
GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCCAAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA
GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera

containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in

Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl2 (anhyd); 0.00130 mg/L $CuSO_4$-$5H_2O$; 0.050 mg/L of $Fe(NO_3)_3$-$9H_2O$; 0.417 mg/L of $FeSO_4$-$7H_2O$; 311.80 mg/L of Kcl; 28.64 mg/L of $MgCl_2$; 48.84 mg/L of $MgSO_4$; 6995.50 mg/L of NaCl; 2400.0 mg/L of $NaHCO_3$; 62.50 mg/L of NaH_2PO_4-H_2O; 71.02 mg/L of Na_2HPO4; .4320 mg/L of $ZnSO_4$-$7H_2O$; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H_2O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H_2O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H_2O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalainine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tryrosine-2Na-$2H_2O$; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B_{12}; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;

0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

Ligand	JAKs tyk2	Jak1	Jak2	Jak3	STATS	GAS(elements) or ISRE
IFN family						
IFN-a/B	+	+	-	-	1,2,3	ISRE
IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
Il-10	+	?	?	-	1,3	
gp130 family						
IL-6 (Pleiotrophic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
Il-11(Pleiotrophic)	?	+	?	?	1,3	
OnM(Pleiotrophic)	?	+	+	?	1,3	
LIF(Pleiotrophic)	?	+	+	?	1,3	
CNTF(Pleiotrophic)	-/+	+	+	?	1,3	
G-CSF(Pleiotrophic)	?	+	?	?	1,3	
IL-12(Pleiotrophic)	+	-	+	+	1,3	
g-C family						
IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
IL-7 (lymphocytes)	-	+	-	+	5	GAS
IL-9 (lymphocytes)	-	+	-	+	5	GAS
IL-13 (lymphocyte)	-	+	?	?	6	GAS
IL-15	?	+	?	+	5	GAS
gp140 family						
IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
IL-5 (myeloid)	-	-	+	-	5	GAS
GM-CSF (myeloid)	-	-	+	-	5	GAS
Growth hormone family						
GH	?	-	+	-	5	
PRL	?	+/-	+	-	1,3,5	
EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
Receptor Tyrosine Kinases						
EGF	?	+	+	-	1,3	GAS (IRF1)
PDGF	?	+	+	-	1,3	
CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:
5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCG
AAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:
5':CTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATG
ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml genticin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1 x 10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

The above protocol may be used in the generation of both transient, as well as, stable transfected cells, which would be apparent to those of skill in the art.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest $2x10e^7$ U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $Na_2HPO_4.7H_2O$, 1 mM $MgCl_2$, and 675 uM $CaCl_2$. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting $1x10^8$ cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of $5x10^5$ cells/ml. Plate 200 ul cells per well in the 96-well plate (or $1x10^5$ cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as $5x10^5$ cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to $1x10^5$ cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF-κB (Nuclear Factor κB) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF-κB regulates the expression of genes involved in immune cell activation, control of apoptosis (NF-κB appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κB is retained in the cytoplasm with I-κB (Inhibitor κB). However, upon stimulation, I- κB is phosphorylated and degraded, causing NF- κB to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κB include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF-κB promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF-kB would be useful in treating

diseases. For example, inhibitors of NF-κB could be used to treat those diseases related to the acute or chronic activation of NF-kB, such as rheumatoid arthritis.

To construct a vector containing the NF-κB promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF-κB binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT: 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF-κB/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF-κB/SV40/SEAP cassette is removed from the above NF-κB/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF-κB/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF-κB/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1,1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 µl of 2.5x dilution buffer into Optiplates containing 35 µl of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 µl Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 µl Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-4 (`Molecular Probes, Inc.; catalog no. F-14202`), used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO_2 incubator for 20 hours.

The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-4 is made in 10% pluronic acid DMSO. To load the cells with fluo-4 , 50 ul of 12 ug/ml fluo-4 is added to each well. The plate is incubated at 37°C in a CO_2 incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10^6 cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-4 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10^6 cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-4 . The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca^{++} concentration.

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members

of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boeheringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg_{2+} (5mM ATP/50mM $MgCl_2$), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM $MgCl_2$, 5 mM $MnCl_2$, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavadin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phospotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,

Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with

specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's

solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P. T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and

variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference. Further, the hard copy of the sequence listing submitted herewith and the corresponding computer readable form are both incorporated herein by reference in their entireties.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 61, line N/A.

B. **IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depositary institution American Type Culture Collection

Address of depositary institution *(including postal code and country)*
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
18 December 1997	209563

C. **ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. **DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indications are not for all designated States)*

EUROPE-
In respect to those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the g rant of the European patent or until the date on which application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28 (4) EPC).

E. **SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☒ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

CANADA

The applicant requests that, until either a Canadian patent has been issued on the basis of an application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the Commissioner of Patents only authorizes the furnishing of a sample of the deposited biological material referred to in the application to an independent expert nominated by the Commissioner, the applicant must, by a written statement, inform the International Bureau accordingly before completion of technical preparations for publication of the international application.

NORWAY

The applicant hereby requests that the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on the list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Regulations), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

UNITED KINGDOM

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for the international publication of the application.

DENMARK

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later that at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person by the applicant in the individual case.

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by a applicant in the individual case.

NETHERLANDS

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapsed, the microorganism shall be made available as provided in the 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever of the two dates occurs earlier.

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

<110> Human Genome Sciences, Inc.

<120> 36 Human Secreted Proteins

<130> PZ022.PCT

<140> Unassigned
<141> 1999-01-06

<150> 60/070,657
<151> 1998-01-07

<150> 60/070,692
<151> 1998-01-07

<150> 60/070,704
<151> 1998-01-07

<150> 60/070,658
<151> 1998-01-07

<160> 168

<170> PatentIn Ver. 2.0

<210> 1
<211> 733
<212> DNA
<213> Homo sapiens

<400> 1
gggatccgga gcccaaatct tctgacaaaa ctcacacatg cccaccgtgc ccagcacctg 60
aattcgaggg tgcaccgtca gtcttcctct tccccccaaa acccaaggac accctcatga 120
tctcccggac tcctgaggtc acatgcgtgg tggtggacgt aagccacgaa gaccctgagg 180
tcaagttcaa ctggtacgtg gacggcgtgg aggtgcataa tgccaagaca aagccgcggg 240
aggagcagta caacagcacg taccgtgtgg tcagcgtcct caccgtcctg caccaggact 300
ggctgaatgg caaggagtac aagtgcaagg tctccaacaa agccctccca acccccatcg 360
agaaaaccat ctccaaagcc aaagggcagc cccgagaacc acaggtgtac accctgcccc 420
catcccggga tgagctgacc aagaaccagg tcagcctgac ctgcctggtc aaaggcttct 480
atccaagcga catcgccgtg gagtgggaga gcaatgggca gccggagaac aactacaaga 540
ccacgcctcc cgtgctggac tccgacggct ccttcttcct ctacagcaag ctcaccgtgg 600
acaagagcag gtggcagcag gggaacgtct tctcatgctc cgtgatgcat gaggctctgc 660
acaaccacta cacgcagaag agcctctccc tgtctccggg taaatgagtg cgacggccgc 720
gactctagag gat 733

<210> 2
<211> 5
<212> PRT
<213> Homo sapiens

<220>
<221> Site
<222> (3)
<223> Xaa equals any of the twenty naturally ocurring L-amino acids

<400> 2
Trp Ser Xaa Trp Ser
1 5

<210> 3
<211> 86
<212> DNA
<213> Homo sapiens

<400> 3
gcgcctcgag atttccccga aatctagatt tccccgaaat gatttccccg aaatgatttc 60
cccgaaatat ctgccatctc aattag 86

<210> 4
<211> 27
<212> DNA
<213> Homo sapiens

<400> 4
gcggcaagct ttttgcaaag cctaggc 27

<210> 5
<211> 271
<212> DNA
<213> Homo sapiens

<400> 5
ctcgagattt ccccgaaatc tagatttccc cgaaatgatt tccccgaaat gatttccccg 60
aaatatctgc catctcaatt agtcagcaac catagtcccg cccctaactc cgcccatccc 120
gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa ttttttttat 180
ttatgcagag gccgaggccg cctcggcctc tgagctattc cagaagtagt gaggaggctt 240
ttttggaggc ctaggctttt gcaaaaagct t 271

<210> 6
<211> 32
<212> DNA
<213> Homo sapiens

<400> 6
gcgctcgagg gatgacagcg atagaacccc gg 32

<210> 7
<211> 31
<212> DNA
<213> Homo sapiens

<400> 7
gcgaagcttc gcgactcccc ggatccgcct c 31

<210> 8
<211> 12
<212> DNA
<213> Homo sapiens

<400> 8
ggggactttc cc 12

<210> 9
<211> 73
<212> DNA
<213> Homo sapiens

<400> 9
gcggcctcga ggggactttc ccggggactt tccggggact ttccgggact ttccatcctg 60
ccatctcaat tag 73

<210> 10
<211> 256
<212> DNA
<213> Homo sapiens

<400> 10
ctcgagggga ctttcccggg gactttccgg ggactttccg ggactttcca tctgccatct 60
caattagtca gcaaccatag tcccgccccct aactccgccc atcccgcccc taactccgcc 120
cagttccgcc cattctccgc cccatggctg actaattttt tttatttatg cagaggccga 180
ggccgcctcg gcctctgagc tattccagaa gtagtgagga ggctttttg gaggcctagg 240
cttttgcaaa aagctt 256

<210> 11
<211> 1713
<212> DNA
<213> Homo sapiens

<400> 11
ggcacgagca cagataaaga taagttttac tgtcatgctg cttttaacat aacagagcaa 60
catcacctag gaaaaaagtt tgtaggagga tttttaatcc atatatttgt cttatggcta 120
gataaagatt tctctgaaaa aaagaagcat gtcaggaatc tctgggtgcc ccttttttcct 180
ctggggactt ctagcattgt tgggcttggc tttggttata tcactgatct tcaatatttc 240
ccactatgtg gaaaagcaac gacaagataa aatgtacagc tactccagtg accacaccag 300
ggttgatgag tattatattg aagacacacc aatttatggt aacttagatg atatgatttc 360
agaaccaatg gatgaaaatt gctatgaaca aatgaaagcc cgaccagaga aatctgtaaa 420
taagatgcag gaagccaccc catctgcaca ggcaaccaat gaaacacaga tgtgctacgc 480
ctcacttgat cacagcgtta aggggaagcg tagaagccca ggaaacagaa tactcatttc 540
tcagacaagg atggagatga gcaactacat gcaatagatg ccagcgtttc taaaccacct 600
tagtagacag tttctccccg aaagccaggc gtagaggaaa acattcatga tgatcccatc 660
gactgtttgg attgatccgt gctaagagaa acctattaac tagctggacc atgatctgtt 720
caatgattgg ctcctattga agatggcttc taagaaaaca agatgcacag aggacacaga 780
aggacttggc agcagggtga tgacctgatc atttgttgat gggatggtgg cttacctctt 840
attcacagct tacacttatg catgccaaat gtaaggccat gaaaatcagt atttcaaata 900
acttaaaaaa tgctttacta ctaaaatgta aaaaattaat gtgctcacct cggcagcaca 960
tatactaaaa attaataaga cccagcttga aaattgagcc tgataacaag attacaaatt 1020
cacaatacct aatacttagg gaaatataaa aatttaagca tgaatgtgtt ctggaacacg 1080
ttagaagaaa aataaaagcc aatgagtttt tttttaattc tcctttctca ccaatgggca 1140
atagcccata attgaaataa atttctgatt gaaaggtata ggaaacatta aaatgcatta 1200
ctaagagaag taatataatt ttcttacaaa gtatttttcc caaagatagc tttactattt 1260
caaaaattgt caaattaatg catgctcctt acaacaaaca aatatcaaaa agagtttagg 1320
aattctacta gccagagata gtcacttgga gaaactttct atatatcctt ctaaatattt 1380
ttctgggcat gcttatgtat gtacatcagt tgtttctttt tattttgaac caaaaatgtg 1440

gtttcttttg tacacattac ttaaactttc tttccagtca acaatatatt gtggatttat 1500
tttcactgtt atatttaact atatataaat acgcatatat tgtaatttta atgtctgctt 1560
agcaccccac tgataaccaa atcacagttt atttaaataa ttttaatgac ttttcaaaaa 1620
caatttattg atgcaaaaag caaggttgag atgacaatgt ttctttcaat aattaaaaaa 1680
tactgcttca ctgtcaaaaa aaaaaaaaaa aaa 1713

<210> 12
<211> 2062
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1674)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (1699)
<223> n equals a,t,g, or c

<400> 12
ggcagagcat taattgatat tttaatatgg atagacattg catagattca aataaattaa 60
aatcaatgat aaatgctaaa tattttatct aaatagtttt tcaagaaaca gttatggaaa 120
tgtgtatatt raatggctct aatgtggagc ttgtggtatt tcaactcagt attcattatt 180
agttgtgtgt ctggaaagat tgtacttact tttcctcttt acactacagt ttgctcttat 240
ggggctctaa actgtttaac tgaagaacct tcgtctgtat tttgattgag cataatttag 300
tattttatga tttccaagat gatgttctta tgtctatcaa gtctatgtat caaatttata 360
acatcattta agaaaaagga atttccacag atacttcagt tgcaattttt tgtttcatgc 420
tactgaaaat acatttgttt ctaggggttg gaatattata gaagatgtag gatgaaagaa 480
aacgatagaa caacgaaaga attctgttta tgaaattaca ggaattgtgt ccactatggt 540
aaagcattgt cattttagta cattttctct tagtagtttg gcattttata ctttaaaact 600
tgttttgctt taaaaattgt ttataatgct taccttcttt ctccagtgcc tttagtcttg 660
atttgatatg tttgtaccct cagttaccct ttctattaca tgtttttgat gttttcatag 720
cctaggaaac atcgattcct ttttaataat tgtcaatctg attatttaaa gaggtaacaa 780
ttatctgtta atgctttgga aaaacaagta gggttgcctt tggaggccag gcttcttart 840
tcattcaaaa atattccttg gatttatgcc atgtwttaag cattttagc ccccagtatt 900
acaactgtga accaaacgga taaggcccta accattttca gcattctctt tggatggggt 960
gggattgggg acttaattaa aatagagata tagaaaaata ggcatctaaa taagataata 1020
agtgtggggt tgaaatgaag catctaacaa tagttgaagt tagaagtaat attttacagt 1080
attgtaacct ctatttaagt ttgggtatta gttacagata gcataaaaaa gccttaattt 1140
ttcactttcc ttgctggcaa aggtacattt atttagactg tccatttaaa gtaatgttta 1200
acataaacat tactgtgaaa aacattccat tacatattcc caagcaaatg agctgcatct 1260
tctttactgt attttacaat ttagtacaac agttttaggc ctcaatctta acatcactgg 1320
tattttaaat ttggcaatga atatgaaatt acttttgact tacagattga ttatattatt 1380
actttgaaaa tgcattaatt tcttagaaaa gtttggagcc tctatctttt tttgagttaa 1440
tacttaaatt ctcattactt atattaatag cctgtactaa gtgaaaatat tatttatgca 1500
agtaaacaag tcactatagg cttttaagac ttttctttaa ttttagattt tgtcatcaaa 1560
gtttaaattt tttacctact gtccacttrr atataattta acagtttgta aagtgaaata 1620
gtyttaagta tgatgtatga tgcaccggca tataatgaaa atggcgtgca caangacact 1680
ttactatggg aactgtacng gaagatttat gaaagcatgt gaaattgcac ctaaaattgt 1740
gttattagtg actataagca gcaatgctaa atttattgta cttgatgaat gaatgtattt 1800
agtcacagtt actttggttt aaatgtataa atgtctttag ggtttttttt taaatgtgtt 1860
tgtaatttgt actattgtgg gggtatactt ggactgcagg ggttattgtc aatgtgtgat 1920
ttgtgttttt attttataga atcatctaat gtgatatacc aatttttata agtgatattt 1980
acataattct aataactgta tatttgacaa cctattaaaa tgttttgcat tggaaaaaaa 2040
aaaaaaaaaa aaaaaactcg ta 2062

<210> 13
<211> 1224
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1205)
<223> n equals a,t,g, or c

<400> 13
ggcacgaggg gactttccac agtcagctgg acgcacactc agcccagtaa aagaggggac 60
ccatcccggg agccccgggg agggcacagc tgcctcctcc cgggctcccc tgccacctgg 120
tgcctacctg ccccctgctc cctgccgggt ccggtcctca ccccatcttc atctggcctt 180
gactctgccc ttgaggggcc taggggtgca gccagcctgc tccgagctcc cctgcagatg 240
gaggaggcca tcctggtccc ctgcgtgctg gggctcctgc tgctgcccat cctggccatg 300
ttgatggcac tgtgtgtgca ctgccacaga ctgccaggct cctacgacag cacatcctca 360
gatagtttgt atccaaaggg gcatccagtt caaacggcct cacacggttg ccccctggcc 420
acctgcctac ccacctgtca cctcctaccc acccctgagc cagccagacc tgctccccat 480
cccaagatcc ccgcagcccc ttgggggctc ccaccggacg ccatcttccc ggcgggattc 540
tgatggtgcc aacagtgtgg cgagctacga aacgaggaac cagcctgtga ggatgcggat 600
gagatgagga cgactatcac aacccaggct acctggtggt gcttcctgac agcaccccgg 660
ccactacact gctgccccat cagctcctgc actcagcacc cctggcatcc gagacagtgc 720
cttctccatg gagtccattg atgattacgt gaacgttccg gagagcgggg agagcgcaga 780
agcgtctctg gatggcagcc gggagtatgt gaatgtgtcc caggaactgc atcctggagc 840
ggctaagact gagcctgccg ccctgagttc ccaggaggca gaggaagtgg aggaagaggg 900
ggctccagat tacgagaatc tgcaggagct gaactgaggg cctgtggagg ccgagtctgt 960
cctggaacca ggcttgcctg ggacggctga gctgggcagc tggaagtggc tctggggtcc 1020
tcacatggcg tcctgccctt gctccagcct gacaacagcc tgagaaatcc ccccgtaact 1080
tattatcact ttggggttcg gcctgtgccc ccgaacgctc tgaccttctg acgcagcctg 1140
agaatgacct gcctggcccc agccctactc tggtaataca ataaaggcct gcgtgtgtct 1200
gtgtnaaaaa aaaaaaaaaa aaaa 1224

<210> 14
<211> 1621
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1596)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (1607)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (1616)
<223> n equals a,t,g, or c

<400> 14
gggcaaccct gccattctcc cctccagtgg aagattctac tttattcagg aaggctggtg 60

gcttgcaggg gctaggctta cacaagccag ccaagattga gaaggtaggg gagtgctttc 120
aatggaggcc atagttgtca tcttttcttc tgagcaggct ctaggacaaa caatgagcaa 180
attgagatgt cctgtattgg gaggatgagg ctgatatgtt tcatcattct cagaatttgt 240
gggctagagc accttttttgg gaatatgggc ctagggraaa agaacggtca cttacctgga 300
cactatggcc atagtcttga gtttttctag tcagcatgtg attggttgta ccttaagttt 360
atgmcaccaa aaatactatt aactctgtta tttttgttct taatcccatc tcaaattagt 420
ttctagagtt agaatcagaa cacaaggatg ctcattttca catggaggga aagttgccaa 480
agcatttaaa aaataacccc aaaagtatat ctagcagtga agcttggtag ataaaggtga 540
actgctaaac agacactggc gtgccacact cctgccaggg cagtctcatt gtgtgtgacg 600
gtgctgcaac actgccagta ttacttgaga tgcagtctct ctcccctgtt ctcagttcct 660
gggctcagcc ctcctccaag gcctgccagt gagtagcwgt ttggaaggca ggccaaggga 720
gatccccaaa gacagttatc agtcttaact tcygctgtct cccgtcaaat acttatacag 780
gcccccatgg gtaataggca agcatgatgg ctgatagaga agtcagtgca tgagttacta 840
tcacatgtcc ccccaacccc tcctgccacc tcccagggct ctggataatt gaatcttcct 900
gagtcccaca gctggagact caaccaggat atagctgtaa atgcccaagt agaatctgac 960
agaataagac agagacactg aaaataaagc cctagaaagg aagaaattgg aagcaaagaa 1020
aaggagaggt gaaaagataa aaagcctcct tcaaggttag gttcaggttc tgttttccat 1080
ttaacctcat gtgccataaa gctgcccagg cacaccagag ccacatcctg aacccgaccc 1140
tccctgacag tgctgctctg ccagtagcaa gccccagatg gaggaagctg ggcccatttc 1200
tggccacttc cacccatttg gagctttgcc agaggagtcg tctatgccaa taatatttct 1260
gcaacagcat attatattat ttgaagatta gtagatcttt ttgggggggk tggggcaggg 1320
gacagtttct atagatgaag aaccagtgtt ggttgtacag ctgttggggg tcatctatcc 1380
catgtgaagc tattcttttt ccaaatcttg ttgtttctgc atttgtgtcc tccaccactc 1440
ccttcttggc tgacatagat atgcctgcca gattktcatc aagggtcata tttcaataaa 1500
aggtgctaag gacaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1560
aaaaaaaaaa aaaaaaaaaa agggcggccg ttttanagga tccaagntta cgtacncgtg 1620
c 1621

<210> 15
<211> 1420
<212> DNA
<213> Homo sapiens

<400> 15
ggcacgagca caagctcaag aggccgcttg cacgcatgtg gacactccat gattctgctt 60
ctatctctct ttcagggcgt gcgaggcagc ctgggctccc ctggaaatcg ggaaaacaag 120
gagaagaagg tcttcatcag cctggtaggc tcccgaggcc ttggctgcag catttccagc 180
ggccccatcc agaagcctgg catctttatc agccatgtga aacctggctc cctgtctgct 240
gaggtgggat tggagatagg ggaccagatt gtcgaagtca atggcgtcga cttctctaac 300
ctggatcaca aggagctgca gctggccggg agctgttcat gacagaccgg gagcggctgg 360
cagaggcgcg gcagcgtgag ctgcagcggc aggagcttct catgcagaag cggctggcga 420
tggagtccaa caagatcctc caggagcagc aggagatgga gcggcaaagg agaaaagaaa 480
ttgcccagaa ggcagcagag gaaaatgaga gataccggaa ggagatggaa cagattgtag 540
aggaggaaga gaagtttaag aagcaatggg aagaagactg gggctcaaag gaacagctac 600
tcttgcctaa aaccatcact gctgaggtac acccagtacc ccttcgcaag ccaaagtatg 660
atcagggagt ggaacctgag ctcgagcccg cagatgacct ggatggaggc acggaggagc 720
agggagagca ggatttccgg aaatatgagg aaggctttga cccctactct atgttcaccc 780
cagagcagat catggggaag gatgtccggc tcctacgcat caagaaggag ggatccttag 840
acctggccct ggaaggcggt gtggactccc ccattgggaa ggtggttgtt tctgctgtgt 900
atgagcgggg agctgctgag cggcatggtg gcattgtgaa aggggacgag atcatggcaa 960
tcaacggcaa gattgtgaca gactacaccc tggctgaggc tgacgctgcc ctgcagaagg 1020
cctggaatca gggcggggac tggatcgacc ttgtggttgc cgtctgcccc ccaaaggagt 1080
atgacgatga gctgaccttc ttctgaagtc caaaagggga aaccaaattc accgttagga 1140
aacagtgagc tccggcccca cctcgtgaac acaaagcctc ggaccagcct tgagagaggc 1200
cacactattc ctttcctctg gcccagtgaa tttggtctct cccagctctg ggggactcct 1260
tccttgaacc ctaataagac cccactggag tctctctctc tccatccctc tcctctgccc 1320
tctgctctaa ttgctgccag gattgtcact ccaaacctta ctctgagctc attaataaaa 1380

taaacagatt tattttccag cttaaaaaaa aaaaaaaaaa 1420

<210> 16
<211> 1035
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (55)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (110)
<223> n equals a,t,g, or c

<400> 16
gagcggataa caaatttcac acaggaaaca gctatgacca tgattacgcc caagntcgaa 60
attaaccttc actaaaggga acaaaagctg gagctccacc gcggtggcgn ccgctctaga 120
actagtggat cccccgggct gcaggaattc ggcacgaggt tacctcctct ctttcagaaa 180
aaagtgttta aatttaataa aaaaatacag acttcctctc tctctgacct gtttctgcac 240
ttctaatttt gtcccattgt tatatctcaa ttctgaaaca agtcccaaac ctttttgtac 300
actcaggctt ttattattta taggtgtctt taatgtggtt tcgctgtttt ttgcttattt 360
ttgtgagcag tgtgactttg acaggtgact ttagaaacat gaagaagcca agcagcctgt 420
gcctctttag acagggcttg atgtctgctt ctgaagttag tggcagcgga agtggagaag 480
gggattgaaa ggtatcttta aattcgraat tatagaaagt aaaaactggt agatgtgagg 540
acagtgggga aactaagatc atagtcgcct aaggttctgt taatacttga gttgaccagg 600
ggggctggtt atgacattga tcatgctaaa ggaaaagatg ccaggaatga gctggggcag 660
agtgaattgg gcagccttcc atcttgacag cacaccaaaa tgtataaatt agcaaaagcc 720
catctttccc taatgccact aagctgtcag tttctggaat tatcatcatt attarattca 780
taatggtttt aatraaggtg tcatccaaac tgacactttg aaaataaagt gagatgatgc 840
ctaaattgga ggcttggaat gaccttagaa aactgctcca ggaaacttga gaatgtccca 900
attacttaaa gaactctgag tcagctacat ggctcattcc attcatttgc tttgcattgg 960
agagatttat ttggattgac acaggttcat gcctcccaga aggctccacc taaaccatca 1020
ctctgctttc tcgag 1035

<210> 17
<211> 858
<212> DNA
<213> Homo sapiens

<400> 17
ggggacatat gtactgtctg tgtggcctat tattacaagc tctcctgaga ctttgtaatg 60
gatacaagac tcaaaagaac catagggagc tgagaatgtg tggtatcatc gcccagggaa 120
aaagccgttg gcagctgcat tgttatccag ggatgaaaag ttagcctgta ttttgcctga 180
gggacgtgat ttgcccttgt tagtacctgt gcctgctttg tcattccagc tgcaggttgg 240
catgctccaa cagcatcatg gacggggcct acacctacgc cgaggtgacc aatgtctcca 300
actgttggat ctgcacaact cttccagcag cagctgggga cagcttgttc tggcacgtgc 360
atccagcttc tatgcagaag tggacatggc tagagacttg gggtcctgca gacaacaggt 420
aggatgcaac acggtgagct ttgaacaggg gttgtcgcag aacccatggc aagcctgccc 480
cctggctgac ttgtagtgtc catgatggat gagcttggct aatgggagaa catgtagtgc 540
ccccattgca ggtaccataa tgtatagagc agcactgtgg taaagtcact gtaagatggt 600
tgcctgctga ggtttgtgca aacaaaacac atgtcaccac aacaagggtg tggtggaaca 660
agcagacttg ccaaggctgg ggccccacgg attttgtgcc ccttgggagt ttatgggtct 720
gtggggacac agaatggcca tctctgccag ccaattggac tggacgttgt acctggaggt 780

ggccctatgt ggctgccact gtgcttccca cctcgtgccg aattcgatat caagcttatc 840
gataccgtcg acctcgag 858

<210> 18
<211> 936
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (893)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (895)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (905)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (911)
<223> n equals a,t,g, or c

<400> 18
gtagataatc tgaatagccc tatatctata gaaacttaat agtgctggga gatataggta 60
ttattatcct cattttacag atgtgaaaat tgaggctcag agaagtaaag tctattgctc 120
aaggtcatgt ggctagaata tggcagagcc atgattcaga tccaggtctt ctgattctta 180
ttccagtgtc ctttctagca taccatgttg cctctaaaga ttgcagctcc ttatttacta 240
gaaaattgtt cctgcccaat ctacatctcc acctcacccc atcttttctt aagcactatg 300
tttgtgtttt tatcagtatt atattcattg tctttggaat acatgttctt gtttgtgttt 360
ggaaaaaaaa tctcttttac cagcttgcac tcggaccaac ttggaaaaaa aaaagcttaa 420
atgtttttgc tatgtacagt ttaaaaatgt gaagtttgta gctttaactt tttgtaagaa 480
aatctaataa cactggctta agtgctgact tgaaatgcta ttttgtaagg tttggatgta 540
agtaatcaat tgaggtcagc agtttgtatg agacatagct tcctccattg cccccactcc 600
ttttttcttt tttaagtttg agatgcttcc tgtgttttta tgttagaatt gttgttctcc 660
ttcttttctt cttcctatac ctcatcacgt ttgttttaaa taaactgtcc tttggaccac 720
aaacccttaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 780
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 840
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aantnaaaaa 900
aaaangaaac naaaaaaaaa aaaaaaaaaa aaaaaa 936

<210> 19
<211> 613
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (602)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (603)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (606)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (607)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (608)
<223> n equals a,t,g, or c

<400> 19
ggcacgaggt catcccttat gtatagtagt taaaggcata aaactgtgac ttttagatat 60
tccacagaac cagacttatt tgatgtggat aataaccaat gatttagcat tttgtttgct 120
tttgttttat tttatccggg ttcatttttt actcttccca tgtacatgaa acaggtggtg 180
gcgtgtagag atcagctgat ccttgtttta tggttaattg aactactttg tatccagggt 240
ttctgcaaat ccaaaagtga tttttcatct aggatctatt cctaacagtc tactccaatc 300
ccactttagt tttccacaat tttaaatctt aatagtgaga attcaaatga aagtcatttc 360
atttgactat tctgatgaca tgattgtggc agaataaatt gggtcttaaa atgccctaga 420
aaatggtaaa tgattaaaaa taatatttta aaattcaacc aaagaaatgg cccattggcc 480
aggtgtggtg gctcacacct gtaatcccag cacttttgga ggctgaggcg ggtggatcac 540
ctgagctcac gagtttgaga ccagcctacc caacatggta aaaccccatc tctacaaaaa 600
annaannnaa aaa 613

<210> 20
<211> 571
<212> DNA
<213> Homo sapiens

<400> 20
ggcacgagtc aaccgtcaaa atgtccaaag aacctctcat tctctggctg atgattgagt 60
tttggtggct ttacctgaca ccagtcactt cagagactgt tgtgacggag gttttgggtc 120
accgggtgac tttgccctgt ctgtactcat cctggtctca caacagcaac agcatgtgct 180
gggggaaaga ccagtgcccc tactccggtt gcaaggaggc gctcatccgc actgatggaa 240
tgagggtgac ctcaagaaag tcagcaaaat atagacttca ggggactatc ccgagaggtg 300
atgtctcctt gaccatctta aaccccagtg aaagtgacag cggtgtgtac tgctgccgca 360
tagaagtgcc tggctggttc aacgatgtaa agataaacgt gcgcctgaat ctacagagag 420
cctcaacaac cacgcacaga acagcaacca ccaccacacg cagaacaaca acaacaagcc 480
ccaccaccac ccgacaaatg acaacaaccc cagctgcact tccaacaacc aaaaaaaaaa 540
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 571

<210> 21
<211> 2024
<212> DNA
<213> Homo sapiens

<220>

<221> SITE
<222> (540)
<223> n equals a,t,g, or c

<400> 21
ggcacgaggt taattgcttt caattgagtc agtaaacctg tgatagataa tttatttaac 60
tggaaaacct aggtacccat aagaaaaaag attcattctc tgtgaaaact gtaggaatct 120
gttgttgttt tcatttgaat atgctctact tctgctctag tatttggttt ggaatatatt 180
ttgtggctct aattactgta tttttaaaaa ccctacctcc attaacagtt ggtaaaggcc 240
ccttttcagg aaagtttgtt gctttttttt ttttttaaa ggaaagctgc tctttgctca 300
gtatagtgtt ttgaaagtga acatagtaac aaatacttta aaaataaagg atacacaatt 360
tatatttgaa aataaaaact ttctgctggt gggattattt atagtccttt atttttaaaa 420
aaaatgtttc ccttttttata ttgctcttga aagtttaatg agcagaatac aatactggtt 480
ataataaaaa tatggtaacc acacagtact cagcctttca atatgttttg ggtcaaactn 540
catttaggca ctagcattta gaaaaatacc aatcacagtg atgctttggt tatttaatat 600
gaaggaaatg gaactaaaac atttatgtca tcaaatttat ttcactcctt tatatttgac 660
tggctggttg atacataatg gttgatgaac atatatttgc ttaaatcact aatagggatg 720
gttgtaaagt aaatagatca ttggttcaac catttaagtg tttttgccag attgtcaatg 780
aaaccatcat actgaccttt tcctccaaaa ttgccaaatt gactgaactg gttgggtgtt 840
tgttaaaatg aavtcaactg tgtgactaaa aagtcagatg gttgccatat tgtttggaat 900
atgttgaatg tcagtgtatg ccttatgtct ttaattgggt atgcaaaaaa atttttactt 960
aagtagatta aaatttttaac ctctagcatg aaaacccagc accaacactg aaagactcca 1020
ttcaggttga agtagcctca aacagtaatt tacttttttga taataggctg ttgtttttct 1080
taaataagct taaaacaatt ctatctgaaa ttggtagcat gggttcactt ggctacaact 1140
gagcaaaata gatgcaactt tcttttaatg gggtgctgcc tctttaggac tgactgtact 1200
atccactgac actggtttgg cagttggtac tgctgaacat ttttatacat gctaccatga 1260
agctatatat gttagtattg aagaagctaa cgggtatact atcatttttg atgtgtgggc 1320
tgattataat ttcctgtatt tcctgtacat tgggatgaaa ctactttagc aaagtccaca 1380
gatcagaaac cagacggtag tttttgaagt tgaaaccagc aaaataagaa aaaataaaaa 1440
atcgatttta attttctgcc tctttcccaa attacccttc ccacttgctc gacaaatcta 1500
tgtaaagcag tttgtttttt catgttttaa ctttaccttg ccctgtgtta tggtactggc 1560
tgacatgtct aagactggat gtgtatattt attatggtgt ctaaaaatca tgaagttcat 1620
cacttttcag gagtatagat aaaatcaaat tggtagtaca tcagagttac ttttcagtgc 1680
accatgacat cactaaaatg agtgctgtaa tgttacaggg ctttcaggtt tgtaaaaaca 1740
taaccataaa ttatattgac gtcagatatg agttgagtat ctataaaata tcacgtgtat 1800
ctcaaaatat tggactgctg tttgactgga tattgctgca taatttttctt ctattgtccc 1860
atatcctttt ggagagagat ttaatgggat ttgaaatgtg caagctgtct aaataagatg 1920
cagtcaaata aagtatggtt aagttgtgtt tgcatttttc ttttagatac agctgtgtgc 1980
attttatgat tgggttgttt tgcttttacc tactaagtaa aaaa 2024

<210> 22
<211> 575
<212> DNA
<213> Homo sapiens

<400> 22
gaattcggca cgagatttgt gaggcctcag ctcacaaaac ctgaatcagg gtccaacagc 60
atctgcctgt ttttattctg caaagctgag atcaagttaa aaggatgatg ataatcactt 120
gtttatctat tcactcagca aacacttatg ggacatttac tgggtgccag gtcttgagct 180
gggcactgag atgagaagct gcctgaggca tgatggtccc atcccactgc cctccagggc 240
tcaaaatact ttatagttta caaagtgttt tcacatgcaa ttttgtgagc tgtgggtgcc 300
attattatcc acattgctta atacatggca gaacctgacc ttaggttgtc caagtccaga 360
ttctaagtcc aagtccagtc cagatcctag agcctgtcca ctctttcctt ccttcctctc 420
tttctttctt gttagttctt tctttttctt tttctctttt ttctttctct ctctttcttt 480
ctttcttcct ttcttctttc tctttttctt ctttctctct ctctctcttt ctttttttca 540
agaccctgtc caaaaaaaaa aaaaaaaaac tcgag 575

<210> 23
<211> 1181
<212> DNA
<213> Homo sapiens

<400> 23
ggcacgagtc gccatctacc ctctagaata aagaaatctt atcattcacc gtctaccctc 60
tagagtaaac aaatcttatc attcgccatc taccctctag aatagagaaa tcttatcatt 120
catcatctac cctccagaat aaagaaatgt tatcatttgc catctaccct ctagaataaa 180
gaaatcttat taaggacatt ttcaaagcct taacagaata tgaatgatta caatattatg 240
ttttacctat acagcatctt ccaagttcta gtttggttgt gccaggccaa acatttgagc 300
cagatttcgg caagatcaag caggagactc tggcgtctgt cgctgattac cttcccgcca 360
tacctggcca ccagcctttc ccatggaccc cacgtgtgtc tccagactct tgggtatgaa 420
tcctgtgaac acactgacct ctgctttcta catgactgac aagcactatg aaatttgcat 480
aaatatgaat agaaaatata ctgtccccac atcccctaaa ataaaactga gtcctcaccc 540
atgggctttt gctggattat actaagaaaa aggcaggtct tacaacacac attcccatag 600
actcacatct cagagaagat tccttccaca ggctcagggt cctgaacaca ctgctccact 660
ctcagggctc ggagacactc tgcaggtggg tcttcacccc atccgggagc cctaattcct 720
tcttcctcgg tttttcatac agtgattttt cccagtactg ctgtctactt tcctctgcca 780
gaaattcctt gctggatcca tggcaatctc ctcaggactc ctatttcata aaatggaaat 840
ggaattcttc atatgcctta aaaagtaaaa taattatcca aacaatataa aatacaaaca 900
cccatctagc aaaatatttc ttagtacatc atgaattcaa atatatatcg tacaaaatga 960
ctgaggattt tattaatttt gagcaatatt gaggttttct cataatatct ttttattgtt 1020
aacttcaaac taaataacaa tcatgtctgg gaataaaatg tataagatag tcatttagaa 1080
aaatttggag atttttcctt ccttgtccag taaatactta acattttacg tgcaatgtat 1140
gtcataaata tgggcattcc cccaaaaaaa aaaaaaaaaa a 1181

<210> 24
<211> 1765
<212> DNA
<213> Homo sapiens

<400> 24
tccagcccgc cacctctgga ctccctcccc tgtatgtaag ccctcaataa aaccccacgt 60
ctcttttgct ggctctgggt ctctttggcg tcttgaacct tatgccttcc ctattgaagt 120
taataggggt tcagcacaac actgggtccc aggcatgcag ccttgtcttt ttatgtttgt 180
cctcatgggc atcatgtggg ccacagggat cttacccaaa atcatgccta gcaggaagag 240
atgcctctct attgatattc cagctgcccc acaagcaggt atgtgcctac tcattctctg 300
acatgtgtcc taggtttctt gactgtaaaa ttccgcttaa taaataatgt attgagcacc 360
tgctatatcc agagctctgc tgggcactct ggaagattag agtttgaaga acacacaatc 420
gttgtcctcc agaagttcat ggtccagtag ggagagaagg tcacaaacag ccagaatcta 480
caagatgaga ggggctatta aagagagctg agaagagttt cagagccact gagaagatac 540
tcattttagt cttggtaacg ggaagctttc tggagttggt ggcatttaag atgaaccaca 600
aatggagcag gaatttaatg aagaggaaag gaccaggacg ttcaaataga ggacacagca 660
gagttcagac cagatgcagg aaagggtaag tggtttggtt tgctgtgagg tcaggtgcat 720
gatggaaaat ggtgagcagt tagatgacaa agataggtta gggcagttat ggtgggagaa 780
accctgtttt tgaagaaaga tacagtcaga tgggcacttc agcagcatta ttctgacagc 840
aatgtgcaaa taggatggag aggcactgaa gaggcaggat gcaaggagaa caattcagac 900
gcagttgaca tcgcttggga gagaatgagg tactatccag gacagcccca agagagtgga 960
gagccctgtc ctgagattct tgggaaaagg tgccccagta tcaacagcat gtgcccataa 1020
gctccctcag cccacttgtc aggctgggtc agcttgtccc ccaccccacc ccagctctct 1080
cctctcctct cagagcaagc cacctgccct actcactgct ccaccatgat acacctgtcc 1140
tgcctcagcc acgtcccccc ctgaccattt ccccctttc aggaaagtgc ttgcagtttg 1200
ctacctccca ttctttgcag cctccttccg caaagctctc accttgtgtc atctttgcct 1260
gtgtgcccta tggagaagga aaacaaagta agcgctgtct tcccagggct gcaggcgtgc 1320
tggcccccta ccttcatgca ctcaaggagc actgactgag catcctgcaa ctctgtcctg 1380

tccccagaag ccatgctttg tgagaagtgg ggtaagcatc acagaagctg gccactggca 1440
ggcgcatgca gggcttcaga cctgctcctg atgtaatctt tgctgcacat tcacctcacc 1500
tgaggacctg tggaaactag cacccaggtc acaggccctg ctaattaaat cagaatgtag 1560
gagccaggca acaataaata gcgctgaaat atccccagct gattccagtg tacagccaag 1620
tttaggaacc gtgtcctggt gtttccctcc tactcttgct aatatgtagc tctacatatc 1680
tgtgtggagt gagactggca gaaaggctgt ctggatgctc cttcctgtat agcttggagg 1740
caaatggact ctatacagct caggg 1765

<210> 25
<211> 891
<212> DNA
<213> Homo sapiens

<400> 25
ggcagagctg gcttgattga aggaaagctt gaaaaggcgc agagccctat acctcatttc 60
ctccatgata aaaggatcca agtgaggccc tgtcacagcc tgtgggtagg ggatgcggcg 120
ggatcctcat tgccatggta ctcaaaggta gaagagcctg gagtttgttg cttctctttg 180
ctattctttc atatcctctt gggcctggtg attaattagc aattctcatt cctctcagcc 240
aaaggcctgc actgggcttt atttgtcttt ttttatttt taagcactgc ctgccagaga 300
tgggcctggg gcctgatgag gaccttagcg ctgctcgttc tccttttctg ttcatgcaca 360
cattcctcca tggggtgggg aaggcaggca tggggtgtgg ccctcggaga agttaggagt 420
cccccagctc aagatacagt ggcaaagacc tagtggtccc ctacccccac ttctctcagt 480
tcctggcatg aggagagaag accctgctct ggtggagctg acaacctttg aggctgggag 540
gagagcagcc tctgggcatc gttcccagtg tccctcacac taaaacggcg tagatggcaa 600
ccccccaccc ccaccccgct gctcaactct tgtgtttgtt gttctgtttg ccccatttat 660
ctgttgctgt ttttgtgttg tcttcccctg tccgcatttt gtaaaatggc ccctggggga 720
gtgtttttgc tggatctgct ccctttcgct ctctcactcc actacttttt ggaacaaagt 780
gatggcagaa tgcggtggtg gtgggggtct tttgtactgt tggattaata aaatgatttt 840
aaaatcccaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 891

<210> 26
<211> 465
<212> DNA
<213> Homo sapiens

<400> 26
actagtggat ccaaagaatt cggcacgagc ccagggaagt taagcaactt gcccactcac 60
tcagtctaat tggtgccggg gcagggacag aactcactca attccactgc tttccccagt 120
ttctggtgag tgatgtgtgc ctggcactgt gtgcacctgg cactgtgtgt agtaggaatg 180
ctgtttcttc tgtctgtaac ctcctcccag ttctgtaaac aaagacaaaa ccatgccctg 240
cccttaaagc ctatagggtt caaatgccac ctctttgatg atgctttccc tataacacca 300
tttgatacct ctcatgggac ttagcaagta attgcaacga gaggcatata tgttagttca 360
atatagcaaa cctttgttga gggcccactg tgtgccaggt atgtctcccc tccctgagga 420
ctggaagcaa aaaaagagaa gagaaaaaaa aaaaaaaaaa aaaaa 465

<210> 27
<211> 783
<212> DNA
<213> Homo sapiens

<400> 27
ggctttttg ggtttatact gggaatttat ttttactaaa ttatttaact tttctaatta 60
tgtaattatg taagctagct tttcatgttt atgtatgtat ggtgtcccct tgtgttattt 120
ttcttcctct tggtttttga attagtgtta aatagaatac tgtctggatt cttaaaatat 180
tttcatttcc atcatggtta taacaaattt gctgcatgcc caaactgaca acagcaatca 240

ctgagggaac aggttttgaa tctttctttt gtgttatgaa gtttatcgtc tctacttgct 300
tgagattttt gttattttgg gggtttgggg gtgctttttg ttttgttttt gccaaatgta 360
acatgaaagc agatgctgca gctttagtct gttatgctga tttagtaaaa aaaaattttt 420
tacatatatt gcttgctttc gatgcttctg tgaaattttt tyctaaagct tttgtgcagc 480
tgtatggtaa aaatatggkg attaattkga agagcttaca ttgaaagaca atgtaatagg 540
aaataaatgt agattgcagt tggtcaagaa ttttgtagag aggataacaa gacttaatta 600
ctgaaaaaca gtaacatagc attttgaaat ataattttta aaatattgat gctttccttt 660
taaatggaaa tttaaawttt atgattaaaa gtttaaacat twatgataat tttcctcatc 720
agttctccca taggaaataa agcatgtgaa agggtaaaaa aaaaaaaaaa aaaaaaactc 780
gag 783

<210> 28
<211> 470
<212> DNA
<213> Homo sapiens

<400> 28
ggcacgagcc agggagatgt ggaccagaaa acctagctct gcaagctacc cagagaggga 60
caagattctc ggtgcctatg tgtaagagtt cacgccagta cacgtataca tcagtgcata 120
tgtgtcagtg tgcgtgtgag cgtgtagagt ggaggggtc actgacacca gccagggcgt 180
tgcataacca cctgacagag caatggtttc cccatggctt ccccttcttg tctcgctttt 240
tcacttactg aattgtctcc gcggtgttgg cacgtctggc cagagcctgg gactgccctc 300
ctcatccttc ccaccaacac cagagcacaa ggccactgcc cgtgactgaa ccacagggat 360
gctgcacacg tccagggatc ctgtcctttc ccttcctttg taagagtgtc ccagcctcct 420
gtcccctaca ctcagaccaa actgttccat ggtggtgact ccagccctga 470

<210> 29
<211> 1321
<212> DNA
<213> Homo sapiens

<400> 29
ggcacgagca ctgagctcta atcactaacc aggcacttgt ccagtcctgg tgggtaggat 60
aagagggtat ggaagtacat ttacctccct gcaggatctt ggaaggaaaa tgattgggac 120
cagagtaatc agtggagaaa aagaactaga caaactgtaa tcagggagtc ttgctgctac 180
aggaggtcat aggaggaaag artcaaagca tgtggtgttg gttggaaata attaaggagg 240
gcttcctgcc caatatggtc taaggatgga aaaactttgt atctcccagt ttgtctttct 300
tttctccata gccctgcctc tactttcctt ccttggaatc aggggttcct tagcccattt 360
gctttctcta ccttggggac cccaggggcc aagcagttct ccatctagtc acaccaaagg 420
caaaaagcct ggctacctcc cccctagcac gtgagtccct actcccctcc cctctgtttc 480
tgcccagctt tgcttatttt ggggatttca aggcagcaga gggtartgak agtaagggga 540
gagcaggmga agsctctgtc ctgtataggc aactgcmtga ctatgcggtg cactagctgt 600
aarccaagat caggtgccca gyccttttg tccattaaca ccccttcttg atctttcaaa 660
ggcagctaat tgctagcaaa tccccccgat tccggtcctt ttccctctat ttctttgtta 720
gaagttttct gtggagctga aaccccagcc tctgtttgac tgggttttca tttaggctta 780
gttgggttct tggagccccc tgtttgttgt tttgtgttgt ttccaatgaa aagcaagttt 840
accctcagag ttatgctttt ccaaagaggc tgatgtcttt gtttttgttt tttttaaatg 900
tttccagggt tctaaagtga aagtgaagtt ggggaagggg gtttgggagt gttaagtaat 960
ccaaggttta gaaacaccca tgagatagtt acccctgatc tccagtccct agctgggggc 1020
tggacagggg gaagggagag aggatttcta ttcaccttta atatattttt acaaaaaaag 1080
caaacaattt aaaaacaagc ccaccgcttc tgtacatgtc taaatatatt tttagaagtg 1140
ggtaggattg tgaatttctg atgcagggcc tttttataaa taggttaggg tagcatcatt 1200
cagacttctc tgttgttttt gtccctgtct ttttcttatg ttgtgttact aatgtaattt 1260
atatttttt tagatcctcc ctttcctata gagataaaag tgatttatct tggcaaatta 1320
t 1321

<210> 30
<211> 620
<212> DNA
<213> Homo sapiens

<400> 30
ggcacgagca ccaaactgcc atattgccgt gagaatgtgt gtctggctta tggtagtgaa 60
tggtcagttt atgcagtggg ctcccaagct catgtctcct tcttggatcc acggcagcca 120
tcatacaacg tcaagtctgt ctgttccagg gagcgaggca gtggaatccg gtcagtgagt 180
ttctacgagc acatcatcac tgtgggaaca gggcagggct ccctgctgtt ctatgacatc 240
cgagctcaga gatttctgga agagaggctc tcagcttgtt atgggtccaa gcccagacta 300
gcaggggaga atctgaaact aaccactggc aaaggctggc tgaatcatga tgaaacctgg 360
aggaattact tttcagacat tgacttcttc cccaatgctg tttacaccca ctgctacgac 420
tcgtctggaa cgaaactctt tgtggcagga ggtcccctcc cttcagggct ccatggaaac 480
tatgctgggc tctggagtta atgacaactc cccaaatgca gagatttcac taacttccaa 540
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 600
aaaaaaaaaa aaaaaaaaaa 620

<210> 31
<211> 1194
<212> DNA
<213> Homo sapiens

<400> 31
gagcccagca acgtgcaagg ggaaagggga caggattctg gatggccatt tgcttcactg 60
ggatgcaaaa cctcttttga gtactagaat cagtatttct tcttccatct ctgctgtacc 120
tgagaagaaa tggccaaacg caccttctct aacttggaga cattcctgat tttcctcctt 180
gtaatgatga gtgccatcac agtggccctt ctcagcctct tgtttatcac cagtgggacc 240
attgaaaacc acaaagattt aggaggccat tttttttcaa ccacccaaag ccctccagcc 300
acccagggct ccacagccgc ccaacgctcc acagccaccc agcattccac agccacccag 360
agctccaaca gccaactcaa acttctccag tgcctttaac cccagagtct cctctatttc 420
agaacttcag tggctaccat attggtgttg gacgagctga ctgcacagga caagtagcag 480
atatcaattt gatgggctat ggcaaatccg gccagaatgc acagggcatc ctcaccaggc 540
tatacagtcg tgccttcatc atggcagaac ctgatgggtc caatcgaaca gtgtttgtca 600
gcatcgacat aggcatggta tcccaaaggc tcaggctgga ggtcctgaac agactgcaga 660
gtaaatatgg ctccctgtac agaagagata atgtcatcct gagtggcact cacactcatt 720
caggtcctgc aggatatttc cagtataccg tgtttgtaat tgccagtgaa ggatttagca 780
atcaaacttt tcagcacatg gtcactggta tcttgaagag cattgacata gcacacacaa 840
atatgaaacc aggcaaaatc ttcatcaata aaggaaatgt ggatggtgtg cagatcaaca 900
gaagtccgta ttcttacctt caaaatccgc agtcagagag agcaaggtat tcttcaaata 960
cagacaagga aatgatagtt ttgaaaatgg tagatttgaa tggagatgac ttgggcctta 1020
tcagtttttc attcagcaag tctgcactag ggacctacta tgagccacgc aatacttcct 1080
tggaatgatg tattccctgg ccttgaaata aggaatctag tacccatgtt tgtgctactg 1140
gaatgaatcc attaaactct ctgagactca aaaaaaaaaa aaaaaaaaaa aaaa 1194

<210> 32
<211> 829
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (32)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (772)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (829)
<223> n equals a,t,g, or c

<400> 32
aggtcaaaaa taataataaa tgttagcatc tnactcagtt ctgtgtgatc cagctcctgc 60
caacccatct gatgagctcc tggttcactc tccttgcttc ctgctttcac ctcctgtggc 120
ccctgtcccg ttcttcccat gtgccaagct cattccagcc cccagacctg tccgctactt 180
ttctcctcca gatcttaggt taggtaacac tccagccccc tctgagatca cctacacacc 240
ttctgttcac ttctgtcatc ctattgctaa attactttgt ctcaaggttc gtaatctctg 300
tgagggagtc ctctctgctg cttttcccaa ggcctgacgc acaccgggca cccagttaac 360
atctttaata tattgggctt tcgtgctacc tgagggcttc tagaaagctc tctacatcca 420
agtgagaggc tggctctcca ggaggtacct gtgctctcag cacaaatgga gcccctctgg 480
gaacaggaac ctgcttcagc cctttagcat acacaggcct gccacatgct ggggacaggg 540
ccacctgagt cccttgcctg gtcccatgtg tggtgccagc tccctccct gatgattgtt 600
ctctctcttt cccaggaccg agatctcact cagcagacag cagcagcccg ggagcctgag 660
ctcaggagga actcttacct ggaaattggg aactgtatgg agactccaaa ctgacttctt 720
tcaaaaaaaa aaaaaaaaaa attttttag ctttgacaaa cacacaaaag tngtaataaa 780
gagagccctc cttgtcaacc caaaatgtga gccccctgtg gaaaaccan 829

<210> 33
<211> 1336
<212> DNA
<213> Homo sapiens

<400> 33
ccacgcgtcc ggagttccac aaaatttgtt agtcatcaaa taatagagtt tttcttaagt 60
gaccacttac atactcatct acaataaacc ccaactaact aatttttcac tttgtgtacc 120
tcatgctaat atccgtggac agtaatgtgc ctgttgtgtt ccttttgctt ttcatccttg 180
tgatcttatg tcacatggaa tgtaaaggcc acatatatat atgtgtgtgt gtgtgtgtgt 240
atatgtatat ttttaagaat atttagtctg gatttcatga aattgacttc tgaaataatt 300
tgcctcaatt ttgtttcctg gtggtttgag aagaaagttc ctgtggtgaa atgaaaaggg 360
gataaaggga agtacttatt ttaaaacata agtaacttgt ggattgttga atactggaaa 420
aagagtgtta cttccccgtt aacctacgcc tcgtgtaatc cttcaggttg gaagtcggat 480
cgcagaccgt gtatatgaca tacccagaaa tttccccctt gctttggatc ttggttgtgg 540
aagaggttac attgcacaat atttgaataa gcttcagtta ttccattgca ggaaactatt 600
ggaaagtttt tccaagctga cattgcagaa aatgctttgt ttgcattggg tgaatgacct 660
tcctagagca cttgagcaga ttcattatat tttaaaacca gatggagtgt ttatcggtgc 720
aatgtttgga ggcgacacac tctatgaact tcggtgttcc ttacagttag cggaaacgga 780
aagggaagga ggattttctc cacacatttc tcctttcact gctgtcaatg acctgggaca 840
tctgcttggg agagctggct ttaatactct gactgtggac actgatgaaa ttcaagttaa 900
ctatcctgga atgtttgaat tgatggaaga tttacaagaa caaaagtcca gaatgttgac 960
ctaattttac aaaacaagct gcatatcagc tgatgaatgc atgagaaatt ttcaaggctt 1020
tcacagtggt cttaaggtat gggtgagagt aactgtgctt ggaatagaaa agccctgctg 1080
catcgagaca caatgctggc agctgcggca gtgtacagag aaatgtacag aaatgaagat 1140
ggttcagtac ctgctacata ccagatctat tacatgatag gatggaaata tcatgagtca 1200
caggcaagac cagctgaaag aggttccgca actgtgtcat ttggagagct aggaaaaata 1260
aacaacctta tgccaccggg gaaaaaatca caataaatat ttattcagtg ttaaaaaaaa 1320
aaaaaaaaaa aaaaaa 1336

<210> 34
<211> 1314
<212> DNA
<213> Homo sapiens

<400> 34
atgcggcagc tcttctatga ccctgacgag tgcgggctga tgaagaaggg gggcttgtac 60
ttcagtgact tctggaataa gctggacgtc ggcgcaatct tgctcttcgt ggcagggctg 120
acctgcaggc tcatcccggc gacgctgtac cccgggcgcg tcatcctctc tctggacttc 180
atcctgttct gcctccggct catgcacatt tttaccatca gtaagacgct ggggcccaag 240
atcatcattg tgaagcggat gatgaaggac gtcttcttct tcctcttcct gctggctgtg 300
tgggtggtgt ccttcggggt ggccaagcag gccatcctca tccacaacga gcgccgggtg 360
gactggctgt tccgagggcc gtctaccact cctacctcac catcttcggg cagatcccgg 420.
gctacatcga cggtgtgaac ttcaacccgg agcactgcag ccccaatggc accgacccct 480
acaagcctaa gtgccccgag agcgacgcga cgcagcagag ccggccttcc ctgagtggct 540
gacggtcctc ctactctgcc tctacctgct cttcaccaac atcctgctgc tcaacctcct 600
catcgccatg ttcaactaca ccttccagca ggtgcaggag cacacggacc agatttggaa 660
gttccagcgc catgacctga tcgaggagta ccacggccgc cccgccgtgc cgcccccgtt 720
gatcctcttc agccacctgc agctcttcat caagagggtg gtcctgaaga ctccggccaa 780
gaggcacaag cagctcaaga acaagctgga gaagaacgag gaggcggccc tgctatcctg 840
ggagatctac ctgaaggaga actacctcca gaaccgacag ttccagcaaa agcagcggcc 900
cgagcagaag atcgaggaca tcagcaataa ggttgacgcc atggtggacc tgctggacct 960
ggacccactg aagaggtcgg gctccatgga gcagaggttg gcctccctgg aggagcaggt 1020
ggcccagaca gcccgagccc tgcactggat cgtgaggacg ctgcgggcca gcggcttcag 1080
ctcggaggcg gacgtcccca ctctggcctc ccagaaggcc gcggaggagc cggatgctga 1140
gccgggaggc aggaagaaga cggaggagcc gggcgacagc taccacgtga atgcccggca 1200
cctcctctac cccaactgcc ctgtcacgcg cttccccgtg cccaacgaga aggtgccctg 1260
ggagacggag ttcctgatct atgacccacc cttttacacg gcagagagga agga 1314

<210> 35
<211> 1264
<212> DNA
<213> Homo sapiens

<400> 35
tgggctggaa cgcgccggaa tctgaggtgt gagtagagcc tgggggagag tggatccagg 60
tgaagggggc agaggactgg gagttttcrt cctcttgaat aagaactcga caacagagtg 120
ggaactttct gtcttgtgat ccattgcctg gtgagtcaca gctcacacca tggatttaac 180
ctgagagctt caacttctgc tttggccctg gagttcccat gccctggtgt cttctaccag 240
ttcttagtgt gttgcactgg agcacagagg acactcgatc gtgcggcgcg cagggcgggg 300
ggccgccgct gcctccccgc gggatggctg gcactgtgct cggagtcggt gcgggcgtgt 360
tcatcttagc cctgctctgg gtggcagtgc tgctgctgtg tgtgctgctg tccagagcct 420
ccggggcggc gaggttctct gtcatttttt attcttcggt gctgtgatca tcacatcagt 480
tctgttgctt ttcccgcgag ctggtgaatt cccagcccca gaagtggaag ttaagattgt 540
ggatgacttt ttcattggcc gctatgtcct gctggctttc cttagtgcca tcttccttgg 600
aggcctcttc ttggttttaa tccattatgt tctggagccg atctatgcca aaccactgca 660
ctcctactga ccactcttca ggaawacgaa aacctgttct ctccttcatt gtgatgacat 720
tgatgagcag gaaggcacta ttcagagcct tgttttgaca gccctcatgc cttaaggtta 780
gaggagtatc tgtccatcac taagacaaat ctctggagtc ctggcttcca gaaacaggat 840
tgccaaattg tccctgtggg gctagattct taccagctta agaaggatat tgctatcttc 900
ttagtacccg taccttagga tttccaactg ttttgaaagg gaaatagtaa cagtgatctg 960
cttagagtgg attttcactc aagtccttag taagtggatt ttggggaaaa aagcacmtgg 1020
gcttctggtt ctttttgata atatataaaa ttattcatta tgaggttgca gttgtttgca 1080
aaggagaggc actcaaattt gaaaggttat tttaatgtga taatttggaa gacttactca 1140
gatgttggtc attgaccact ctgtgcatat atttctgcag agctctgtga aggcaatgag 1200
tgtcacttcc ctctgctcta ataaagcaat aaataataaa aaaaaaaaaa aaaaaaaact 1260
cgag 1264

<210> 36
<211> 688
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (607)
<223> n equals a,t,g, or c

<400> 36
cccacgcgtc cgggtcccga tgagcctcct gttgcctccg ctggcgctgc tgctgcttct 60
cgcggcgctt gtggccccag ccacagccgc cactgcctac cggccggact ggaaccgtct 120
gagcggccta acccgcgccc gggtagagac ctgcggggga tgacagctga accgcctaaa 180
ggaggtgaag gctttcgtca cgcaggacat tccattctat cacaacctgg tgatgaaaca 240
cctccctggg gccgaccctg agctcgtgct gctgggccgc cgctacgagg aactagagcg 300
catcccactc agtgaaatga cccgcgaaga gatcaatgcg ctagtgcagg agctcggctt 360
taccgcaagg cggcgcccga cgcgcaggtg ccccccgagt acgtgtgggc gcccgcgaag 420
cccccagagg aaacttcgga ccacgctgac ctgtaggtcc gggggcgcgg cggagctggg 480
acctacctgc ctgagtcctg gagacagaat gaagcgctca gcatcccggg aatacttctc 540
ttgctgagag ccgatgcccg tccccgggcc agcagggatg gggttgggga ggttctccca 600
accccanttt cttccttccc cagctccact aaattccctc ctgccttaaa aaaaaaaaaa 660
aaaaaaaaaa aaaaaaaaaa aaaaaaaa 688

<210> 37
<211> 1516
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (34)
<223> n equals a,t,g, or c

<400> 37
ttcacgtgca cactgatcac aacgtcacgc ctgncagggc accggtccgg gaattcccgg 60
gtcgacccac gcgtccgcag gagaggtgtc tgtgcgtcct gcacccacat ctttctctgt 120
cccctccttg ccctgtctgg aggctgctag actcctatct tctgaattct atagtgcctg 180
ggtctcagcg cagtgccgat ggtggcccgt ccttgtggtt cctctctact tggggaaatc 240
aggtgcagcg gccatggcta cagcaagacc cccctggatg tgggtgctct gtgctctgat 300
cacagccttg cttctggggg tcacagagca tgttctcgcc aacaatgatg tttcctgtga 360
ccacccctct aacaccgtgc cctctgggag caaccaggac ctgggagctg gggccgggga 420
agacgcccgg tcggatgaca gcagcagccg catcatcaat ggatccgact gcgatatgca 480
cacccagccg tggcaggccg cgctgttgct aaggcccaac cagctctact gcggggcggt 540
gttggtgcat ccacagtggc tgctcacggc cgcccactgc aggaagaaag ttttcagagt 600
ccgtctcggc cactactccc tgtcaccagt ttatgaatct gggcagcaga tgttccaggg 660
ggtcaaatcc atcccccacc ctggctactc ccaccctggc cactctaacg acctcatgct 720
catcaaactg aacagaagaa ttcgtcccac taaagatgtc agacccatca acgtctcctc 780
tcattgtccc tctgctggga caaagtgctt ggtgtctggc tgggggacaa ccaagagccc 840
ccaagtgcac ttccctaagg tcctccagtg cttgaatatc agcgtgctaa gtcagaaaag 900
gtgcgaggat gcttacccga gacagataga tgacaccatg ttctgcgccg gtgacaaagc 960
aggtagagac tcctgccagg gtgattctgg ggggcctgtg gtctgcaatg gctccctgca 1020
gggactcgtg tcctggggag attacccttg tgcccggccc aacagaccgg gtgtctacac 1080
gaacctctgc aagttcacca agtggatcca ggaaaccatc caggccaact cctgagtcat 1140
cccaggactc agcacaccgg catccccacc tgctgcaggg acagccctga cactcctttc 1200

agaccctcat tccttcccag agatgttgag aatgttcatc tctccagccc ctgaccccat 1260
gtctcctgga ctcagggtct gcttcccca cattgggctg accgtgtctc tctagttgaa 1320
ccctgggaac aatttccaaa actgtccagg gcgggggttg cgtctcaatc tccctggggc 1380
actttcatcc tcaagctcag ggcccatccc ttctctgcag ctctgaccca aatttagtcc 1440
cagaaataaa ctgagaagtg gaatcttaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1500
aaaaaagggc ggccgc 1516

<210> 38
<211> 1267
<212> DNA
<213> Homo sapiens

<400> 38
ggcacgagct gcaggggcgg ggcggcgcca agcgcaggga gcccggctga gtggcagccc 60
agattgaaga tggatacgtg acaatcccag ggaccgctgc actgacttca tttccttaga 120
caagacacag tgtagggccc ggcccgtgtt ggccccagga ctcctttgga atatagctgt 180
ggacaatgaa tcctgcgagc gatgggggca catcagagag cattttgac ctggactatg 240
catcctgggg gatccgctcc acgctgatgg tcgctggctt tgtcttctac ttgggcgtct 300
ttgtggtctg ccaccagctg tcctcttccc tgaatgccac ttaccgttct ttggtggcca 360
gagagaaggt cttctgggac ctggcggcca cgcgtgcagt ctttggtgtt cagagcacag 420
ccgcagctgt gggctctgct gggggaccct gtgctgcatg ccgacaaggc gcgtggccag 480
cagaactggt gctggtttca catcacgaca gcaacgggat tcttttgctt tgaaaatgtt 540
gcagtccacc tgtccaactt gatcttccgg acatttgact tgtttctggt tatccaccat 600
ctctttgcct ttcttgggtt tcttggctgc ttggtcaatc tccaagctgg ccactatcta 660
gctatgacca cgttgctcct ggagatgagc acgcccttta cctgcgtttc ctggatgctc 720
ttaaaggcgg gctggtccga gtctctgttt tggaagctca accagtggct gatgattcac 780
atgtttcact gccgcatggt tctaacctac cacatgtggt gggtgtgttt ctggcactgg 840
gacggcctgg tcagcagcct gtatctgcct catttgacac tgttccttgt cggactggct 900
ctgcttacgc taatcattaa tccatattgg acccataaga agactcagca gcttctcaat 960
ccggtggact ggaacttcgc acagccagaa gccaagagca ggccagaagg caacgggcag 1020
ctgctgcgga agaagaggcc atagctgctc cagccggggc tccggggcgg cagcagagct 1080
ggcacaccga ttctgggaag ccccgcgaat gatggctttt gaattaatga ggcagtgaat 1140
gttttgtgtt tacttctaag ggaaatacta actttctttc gcattagtat taattttgaa 1200
gtagctacaa agtattttta agaaattata attttatgac tgtcaaaaaa aaaaaaaaaa 1260
aaaaaaa 1267

<210> 39
<211> 2203
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1246)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (1846)
<223> n equals a,t,g, or c

<400> 39
ggcacgagcg gaggcgcgcg ggccagcctg ggccccagcc cacaccttca ccagggccca 60
ggagccacca tgtggcgatg tccactgggg ctactgctgt tgctgccgct ggctggccac 120
ttggctctgg gtgcccagca gggtcgtggg cgccgggagc tagcaccggg tctgcacctg 180
cggggcatcc gggacgcggg aggccggtac tgccaggagc aggacctgtg ctgccgcggc 240

cgtgccgacg actgtgccct gccctacctg ggcgccatct gttactgtga cctcttctgc 300
aaccgcacgg tctccgactg ctgccctgac ttctgggact tctgcctcgg cgtgccaccc 360
ccttttcccc cgatccaagg atgtatgcat ggaggtcgta tctatccagt cttgggaacg 420
tactgggaca actgtaaccg ttgcacctgc caggaraaca ggcagtggca gtgtgaccaa 480
gaaccatgcc tggtggatcc agacatgatc aaagccatca accagggcaa ctatggctgg 540
caggctggga rccacagcgc cttctggggc atgaccctgg atgagggcat tcgstaccgc 600
ctgggcacca tccgcccatc ttcctcggtc atgaacatgc atgaaattta tacagtgctg 660
aacccagggg aggtgcttcc cacagccttc gaggcctctg arartkkycc amcctgwttc 720
awsagcctct wgaccaaggc aactgtgcag gctcctgggc cttctccaca gcagctgtgg 780
catccgatcg tgtctcaatc cattctctgg gacacatgac gcctgtcctg tcgccccaga 840
acctgctgtc ttgtgacacc caccagcagc agggctgccg cggtgggcgt ctcgatggtg 900
cctggtggtt cctgcgtcgc cgaggggtgg tgtctgacca ctgctacccc ttctcgggcc 960
gtgaacgaga cgaggctggc cctgcgcccc cctgtatgat gcacagccga gccatgggtc 1020
ggggcaagcg ccaggccact gcccactgcc ccaacagcta tgttaataac aatgacatct 1080
accaggtcac tcctgtctac cgcctcggct ccaacgacaa ggagatcatg aaggagctga 1140
tggagaatgg ccctgttcca agccctcatg gaggtgcatg aggacttctt cctataacaa 1200
gggargcatc tacagccaca cgccagtgag ccttgggagg ccaganagat accgccggca 1260
tgggacccac tcagtcaaga tcacaggatg gggagaggag acgctgccag atggaatgac 1320
gctcaaatac tggactgcgg ccaactcctg gggcccagcc tggggcgaga ggggccactt 1380
ccgcatcgtg cgcggcgtca atgagtgcga catcgagagc ttcgtgctgg gcgtctgggg 1440
ccgcgtgggc atggaggaca tgggtcatca ctgaggctgc gggcaccacg cggggtccgg 1500
cctgggatcc aggctaaggg ccggcggaag aggccccaat ggggcggtga ccccagcctc 1560
gcccgacaga gcccggggcg caggcgggcg ccagggcgct aatcccggcg cgggttccgc 1620
tgacgcagcg ccccgcctgg gagccgcggg caggcgagac tggcggagcc cccagacctc 1680
ccagtgggga cggggcaggg cctggcctgg gaaaagcaca gctgcagatc ccaggcctct 1740
ggcgccccca ctcaagacta ccaaagccag gacactcaag tctccagccc cactacccca 1800
ccccactcct gtattctttt ttttttttt ttagacaggg tcttgnctcc gttgccccag 1860
gttggagtgc agtggcccat cagggctcac tgtaacctcc gactcctggg ttcaagtgac 1920
cctcccacct cagcctctca agtagctggg actatcaggt gcaccaccac acctggctaa 1980
tttttgtatt ttttgtaaag agggggtct cactgtgttg cccaggctgg tctcgaactc 2040
ctgggctcaa gcggtccacc tgcctccgcc tcccaaagtg actgggattg ccaggcatga 2100
gccactgcac ccagtccctg tattcttatt cttcagatat ttatttttct tttcactgtt 2160
ttaaaataaa accaaagtat tgataaaaaa aaaaaaaaaa aaa 2203

<210> 40
<211> 1726
<212> DNA
<213> Homo sapiens

<400> 40
cgtctgatta aggtaccttt tgggaaatta aggttctata gaaattactg ggctcaatct 60
agtgatacaa atatgtgttg tttgatttat caacacatta caaaccttaa ctttggagtt 120
ttaatatctg gttatcttta atatctggtt atcttctttc tgaagtgtat gtacacaaaa 180
ttgatgctaa ataaggtctt gttgttttgg caaatagtga aatgcaaggt attggtagat 240
cagtactgtt ataactttgg tgcaaagttg ctgcatgcag attggctgtg ggaccttgtt 300
catttttga gaactaatgt agagtttgaa aaaacaccgt aagcctgcat tccagaagtt 360
ctggtatgga tagtgtgagc ccagggaatg tgcttagata aaagatcatt taacaaatag 420
gttttgcatt tttttagcaa tcaggctttg tgctgaatat tagagtggtt gtttcagaga 480
gtttgcagca attaggcttt attggtgcac taaggagaag cagagaggag aagcaattct 540
tggtaacttc cttggaagtt gcagctaact ctgaaaagtc tgggttgaac taggtaagta 600
actaattcct agaatcaata aactttgcag gagtccgttt gattgtacat gtagctccct 660
ggaattgcta ttggtcccta aatcatcagt ttgtaatgct ggttttcaaa cttgagtgca 720
catcaagttt tggaggactt gttagaatac agattgctgg gctcaccccc agagtttctg 780
atctggtagg tctggagcgg gacctggtag attgcatttc taaaaagcat ccaggtaata 840
ctgctgctgt ttgggaaagt acctttgaga tcactggctt acagcaatct caaggtgttt 900
ggattttggg caggggtgct gtgcaggcgt tgctgggatc tcttcacagc actccactgc 960
atagaggtga gcctccagat gttttcattc attcaacaaa tatatgtacc tattgtgtgc 1020

tgggcactgc ttaagttgcg aggggatatt gtgaagaaag taagcaaaac ccctttgttt 1080
gtagaatttc agtgagcata gtcctgggtt aacctgacaa cagtcctact gtttattgat 1140
gcttataggt gagcctattt ctctttctag ctttcttcca cttaatttac tttcttttgg 1200
aattcttgaa tttaataata ataatattga tgttattagt catcactata acttttttatt 1260
gagtatgtat tttatgtcag acacagtgtg gctaagtgct ttacatacat tatctcatct 1320
aatccttaga aaaaaccctg gtgtattagt cttaatttaa aagatgtact ttggaaaggt 1380
tagtagttta cccaagatta tgcagctagt taaaagtggt gctggggctg ggcttggtgg 1440
ctcacacctg taatcgcagt gctttgggag tctgaggcag gaggatcgct tgacaccagg 1500
agtttgagac taacctggga aacatagcaa gaccccatct ttacaaaaaa taaaaaaatt 1560
agccaggagt gggggtgcac acctgtggtc ccagctactt gggaggctga ggcaggagga 1620
ttgcttgagc cccagaggtt gaggctgcag tgagccatga ttatgccact gcacatctgt 1680
ctgggtgaca gagcaagatc ctgtctccaa aaaaaaaaaa aaaaaa 1726

<210> 41
<211> 1102
<212> DNA
<213> Homo sapiens

<400> 41
ggcacgagtt tttatgctgg ttggtctctt gcagggaggt ggtgctgtcc agagagtatc 60
agctgtggta gtatagagag gaacaggtgg tggatggggc cctagaactc ccaggagtat 120
atgccctttg tcttcagtta ccagggtagg tagggaaggc tgttgggtca gggcagggct 180
aggcatgtct gagctcagac tcttcttggg caggtcttgc tgtggctgct gtgggggatg 240
ggggtgaggt tcccaggtca atggtgttat gttcctagga agtttatgtt tttctctcct 300
gtgtcatgca ggttgtcagg gaagtggggg aaagccagca gtcacaggcc tcacccagct 360
cccacataat ccaaagggct ggtttcactc ccaccatgcc ccccgacccc aatagcaccg 420
agtctgtttt aggcagtgga caagcagggc tgagaacttg ccccaggcta cctgcctccc 480
agctgtgaaa gcaagtttgg ctttccttct tcccctgcct gtagagtctg cacactggat 540
tcaggccctc ccccaagttc tggccaggag acttctcgat tagttcaaat tgttgccaag 600
ttcagctgga gatttccttc tcctgtggcc tttcccagtg cctctggcca ccctaccgaa 660
ggacaatgtg aggccaggca gaaatagttt gctaggggac ccagagagct cacagggctt 720
ttcctctacc cctgtatttt gaggggtaaa ggaaaatctg ctgtttcctc taccctgta 780
ttttgcttgg ttctctcaat ttactcagct ccaggtaaga tcagcatcat ctcccgtaat 840
ctagaccttc agtttcctta gtggggtgtg tgttcaggag cagacgatct ccctttccca 900
cttccacggt ttgggcactc acagtatttg gatcatctcc tgggtccagc aggagcaatc 960
tgcttccttc agaaggtctg tgggtccccc ccggcttcct gatttattcc tgccgtcatt 1020
gaagacagag gcgacttcct ccttggaaac tttaggcgcc actgctcagt ccgaaaaagc 1080
ccattccatt ttctactaat tg 1102

<210> 42
<211> 1397
<212> DNA
<213> Homo sapiens

<400> 42
tcgacccacg cgtccgctga attgcggccg tatgcgcggc tctgtggagt gcacctgggg 60
ttgggggcac tgtgccccca gccccctgct cctttggact ctacttctgt ttgcagcccc 120
atttggcctg ctgggggaga agacccgcca gctgcttgag tttgacagca ccaacgtgtc 180
cgatacggca gcaaagcctt tgggaagacc atatcctcca tactccttgg ccgatttctc 240
ttggaacaac atcactgatt cattggatcc tgccaccctg agtgccacat ttcaaggcca 300
ccccatgaac gaccctacca ggacttttgc caatggcagc ctggccttca gggtccaggc 360
cttttccagg tccagccgac cagcccaacc ccctcgcctc ctgcacacag cagacacctg 420
tcagctagag gtggccctga ttggagcctc tccccgggga aaccgttccc tgtttgggct 480
ggaggtagcc acattgggcc agggccctga ctgcccctca atgcaggagc agcactccak 540
cgaacgatga atatgcaccg gccgtcttcc agttggacca gctactgtgg ggctccctcc 600
catcaggctt tgcacagtgg cgaccagtgg cttactccca gaagccgggg ggccgagaat 660

cagccctgcc ctgccaagct tcccctcttc atcctgcctt agcatactct cttccccagt 720
cacccattgt ccgagccttc tttgggtccc agaataactt ctgtgccttc aatctgacgt 780
tcggggcttc cacaggccct ggctattggg accaacacta cctcagctgg tcgatgctcc 840
tgggtgtggg cttccctcca gtggacggct tgtccccact agtcctgggc atcatggcag 900
tggccctggg tgccccaggg ctcatgctgc taggggggcgg cttggttctg ctgctgcacc 960
acaagaagta ctcagagtac cagtccataa attaaggccc gctctctgga gggaaggaca 1020
ttactgaacc tgtcttgctg tgcctcgaaa ctctggaggt tggagcatca agttccagcc 1080
ggcccccttca ctcccccatc ttgcttttct gtggaacctc agaggccagc ctcgacttcc 1140
tggagacccc caggtggggc ttccttcata ctttgttggg ggactttgga ggcgggcagg 1200
ggacagggct attgataagg tccccttggt gttgccttct tgcatctcca cacatttccc 1260
ttggatggga cttgcaggcc taaatgagag gcattctgac tggttggctg ccctggaagg 1320
caagaaaata gatttatttt ttttcamaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1380
aaaaaaaggg cggccgc 1397.

<210> 43
<211> 1739
<212> DNA
<213> Homo sapiens

<400> 43
ggcacgagag atcctcagga tatctttagc caaaggaaaa gctccgcatt cccacctggg 60
gggaaagctg gattgccatg ggcacgaata gtggtgcaga gtccctggcc atcctgaata 120
tccagaatgg tgtttctgaa gttcttctgc atgagtttct tctgccacct gtgtcaaggc 180
tacttcgatg gccccctcta cccagagatg tccaatggga ctctgcacca ctacttcgtg 240
cccgatgggg actatgagga gaacgatgac cccgagaagt gccagctgct cttcagggtg 300
agtgaccaca ggcgctgctc ccaggggggag gggagccagg ttggcagcct gctgagcctc 360
accctgcggg aggagttcac cgtgctgggc caccaggtgg aaggatgctg ggcgcgtgct 420
ggagggcatc agcaaaagca tctcctacga cctagacggg gaagagagct atggcaagta 480
cctgcggcgg gagtcccacc agatcggggg atgcctactc caactcggac aaatccctca 540
ctgagctgga gagcaagttc aagcagggcc aggaacagga cagccggcag gagagcaggc 600
tcaacgagga ctttctggga atgctggtcc acaccaggtc cctgctgaag gagacactgg 660
acatctctgt ggggctcagg gacaaatacg agctgctggc cctcaccatt aggaccatgg 720
gacccgacta gtcggctgaa aaatgattat cttaaagtat aggtggaagg atacaaatgc 780
ttagaaagag ggaatcaaat cagccccgtt ttggaaggtg ggggacagaa aatggggcta 840
catttccccc atacctacta ttttttttata tcccgatttg cactttgaga atacatctaa 900
ggtcatcttt caaaagagaa aaattggaca cttgagtgac tttgttttta gttttgtttt 960
tgaacattat ttatgtgatt gttatggaat tgtcacctgg aaagaacaat tttaagcaat 1020
gtcatttcta gatgggtttc taattctgca gagacacccg tttcagccac atctaaaaga 1080
gcacagttta tgtggtgcgg aattaaactt ccccatcctg cagattatgt ggaaataccc 1140
aaagataata gtgcatagct cctttcagcc tctagccttc actcctgggc tccaaaagct 1200
atcccagttg cctgtttttc aaatgaggtt caaggtgctg ctttgcatgc ctgccaaccc 1260
atggaagttg tttcttactt cttttctctc ttatttatta accatggtct gagagttgtt 1320
tttgttctat gtaacagtat tgccacaaaa ctataggcaa atcgtgtttg cagggagatt 1380
tctgatgcct ctgtgggtgt gtgtaagtta aagtggccac atttaagaag gccaagcttt 1440
gtagtggttg cacagtcaca ctgatatgct gatttgctct ttctcattgt atgtctatgc 1500
tttgtcatca gtgctatagt aaattacaaa gaaataggta gattgtatga acatacccac 1560
aaatgcctat gatttaggtt accaatgtat tctttctcat ttggggtttt gcttctgtct 1620
gtctgtttat tggaaacttg tacttcaagt agggggaatc ctaattctaa taactcctta 1680
gctaagtttt attattcagg caataaacat gttttcatgt aaaaaaaaaa aaaaaaaaa 1739

<210> 44
<211> 3061
<212> DNA
<213> Homo sapiens

<220>

<221> SITE
<222> (2755)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2849)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2919)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2983)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2987)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2998)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (3027)
<223> n equals a,t,g, or c

<400> 44
gtgtgggggc caccttcggt ggcggccgct ctagaactag tggatccccc gggctgcagg 60
gaattcggca cgagcaacat tgaagaccgg gacgagcttg cctaccacat cagcatcatg 120
ttctatataa taggaggtgt ggccactctc ctcctcatcc ttgtcatcat tgtgttcaag 180
gagaaaccta aatatccccc cagcagggcc caatccctga gctatgcctt gacctctcct 240
gatgcctcat acttaggttc catcgcccgg ctcttcaaaa atctcaactt tgtgctgctt 300
gtcatcacct atggtctgaa tgctggtgct ttttatgcct tgtccactct tctgaatcgc 360
atggtgatct ggcactaccc gggggaagaa gtgaatgctg gaagaattgg cctgacgatc 420
gtcattgcag gaatgcttgg ggctgtgatc tcaggaatct ggctggatag gtccaaaacc 480
tacaaagaga caaccctggt agtctatatc atgacactgg tgggcatggt ggtgtacacg 540
tttaccttga acctgggaca cctgtgggta gtgttcatca ctgctggcac aatgggcttc 600
tttatgactg gctatctccc actgggattt gagtttgctg tggagctcac gtacccagaa 660
tcagaaggca tctcctccgg cctcctcaac atatctgcac aggtatttgg gatcatcttt 720
accatctccc agggccagat tattgacaac tatggaacca agcctgggaa catcttcctg 780
tgtgtgttcc ttactcttgg agcagccctc actgcattca ttaaggcaga tctccggaga 840
cagaaagcaa acaaagaaac tcttgagaac aaactccaag aggaggagga ggagagcaac 900
accagcaaag tgcccactgc tgtgtcagag gatcatctct gagaggaagg tggtgacaac 960
tcagggaaca cgaacacccc acctttttcct tcagcacagc tctcaccgcc agcacaaagg 1020
gcttcgctag agatgttttt ggagggaatc agtgggacta tttgtggcat ggatggccta 1080
ttcctcctag aacccacgta agagcttgga tgatttagtt ggagaaaatt gcacctatca 1140
ccaaatgcaa atttgattcc cacctccacc ccctttttagg ttatgggagt tggtgttggg 1200
acagggtggc agagaatatt ggagtcaatc ctagcttggt ctcttgcctt ccctcttttc 1260
ctccatccat cgtggacaat gcctgcaaaa ttttcacagg aagaaagcct attcaggata 1320
ttaacttgaa atttccagtg tcctaagagc ctctcatgaa gcccagttct aataagtggc 1380

aagctgctct gccggggtca tctcctgggt catcggactg attgctcaag ttctgcagga 1440
gaggaagcac cattagaaca actccatcag aacagctcca ccgggacttg tgggcctaaa 1500
ttttcctggc ctaacgggtc tgtctccaaa ccctctttcc taagagctga gcaaaccaac 1560
cataataaac ttgacaaaag actttgttgt ggccatgaca gagataccga ctcaggaggg 1620
ctacctacct aggtgtgatc atgctggggg ctaccttctg agtatatttg tgaaagcaca 1680
tatttgggaa ctctggtagc ttgagttggg aatgggaagg ttctttttta cagaagtact 1740
tccccaggga cttctgtgtg tcacagtcac ctctgatgcc tttatcttga tgttgcattg 1800
ggaatctcag ccatcagccc aagtgcttgt tttattccaa ggcagggtaa tccccgtcaa 1860
cttactctaa cctttgctga aaactaatct tgattcattc tactctgaaa atccaaaggt 1920
gcttctgaga gataagaggg aaggggtaga aggaaaggtg ccccttgaaa tgggaattga 1980
gcctgttaga attaaaagct tatctcacct ctgctgggga cagtatttgc accaccaacc 2040
cctctcctca cctgctttga gcgataatct ttatcagata ttctaaactt aaagggattc 2100
cctttaaacc aactcaagct gatctttcct atctagcctg ctgtttggct gtactcatgg 2160
gctttggtaa tatctcctaa aaatgaggtt ttggtaattt ttcctatgca ttgggcaact 2220
gtgatcgtga ccactgtgct gtcttgctcc agccactgcc ctggcctcag catatcaggg 2280
cagcctgtgc tggctgcaat actgtggtgc ttgggccact gcctgagagg agccaggttt 2340
gtgtgtgtct gcatgtgtgt gtgtgtgtgt ttgtacagat tcaagcaatg gatgcaagga 2400
acatgctgta tgtaatagaa gaaagaagtc cacgttttcg gcagaagtag tgagtcagtg 2460
tggaagagag gtgagggtgt gctttacttt ttgataaaga gaaagatgtt tactcataaa 2520
cccttcaaaa ggtattaaca aaatgtttac caaacctatt gctttatttt aaaaacataa 2580
tttgtgtttt ctatttgtaa gatctgacat ttcgaggcaa taaaaacttc tcagaaaaaa 2640
aaaaaaaaaa aaaaactcga ggggggggccc ggtacccaat tcgccctata gtgagtcgta 2700
ttacaattca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgtnaccca 2760
acttaatcgc cttgcagcac atccccctttt cgccagctgg cgtaatagcg aagaggcccg 2820
caccgatcgc ccttcccaac agttgcgcng cctgaatggg cgaatggcaa attgtaagcg 2880
ttaatatttt ggttaaaatt cgcgttaaat ttttgtttna atcagctcat tttttaacca 2940
ataggccgaa atcggcaaaa tcccttatta atcaaaagaa tanaccnaaa tagggttnaa 3000
tgttgttcca tttggaacaa gagtccncta ttaaagacgt ggactccacg tcaaagggcg 3060
a 3061

<210> 45
<211> 974
<212> DNA
<213> Homo sapiens

<400> 45
caggagtaaa gaactttatg agttcatgag aacctaaggc tcagtatttg aaaattactg 60
acttatgaga aagcaggcat gtaaataaaa aataaaaaat gttggcccta gattttgata 120
tgtgtgtggt gtgtggtgta ggagaggccc tgatatttac ctgtaagtgt tagagttgta 180
tgaaaaaggt ggcaagattg agtagcttag ggcatgtggt gtggaggctg tatgctagag 240
ttttggcatt aataacttgt attttctggg ttttggcatt aataatttgt attttcactc 300
cccaaatatt tttcaagcat ctacttcatg ccagaccttg ttctagatac cggagataca 360
acagcaaaaa tacagatctt gcccttatga agcttaaatt gttgaggcag gcagacagtg 420
ataaataaat acatagaatg ttgggaaaga aaataagagg atttgagagg gtggaatggg 480
gaagaaagga ttcactgata agatgccatt cgagctaaga cctgaaaagg tgatctctaa 540
ggtgaggaaa gagctttcta cacagaagga acagctgggg gaagggagca cgcttggaat 600
atttaaggaa tatcaaggag ggaaaagtgg ctagagtaga ggaagagaat ggaagaagtc 660
atgtcaaaca ggtactaatg gaagaagtca tgtcagacag ggtcttgccc attgtaagga 720
ctttggcctt atacctcagc aagctgagca gccgtcggaa tgttttaagc aaaagagtga 780
caccatcttt aaaagggacc ccttgtaagg attcagaaca gacttggagg gaaaacaagt 840
agaagcagca gggggactag ttaggaggct gaggtgggag gattgcttgg gcctgggagg 900
ttgaagctgc agtgagtcat gatcactcca ctgcactcca gcctgggcga cagaaggaga 960
ccctgtctca gaaa 974

<210> 46
<211> 1446

<212> DNA
<213> Homo sapiens

<400> 46
ccacgcgtcc gagcaaccgc agcttctagt atccagactc cagcgccgcc ccgggcgcgg 60
accccaaccc cgacccagag cttctccagc ggcggcgcac gagcagggct ccccgcctta 120
acttcctccg cggggcccag ccaccttcgg gagtccgggt tgcccacctg caaactctcc 180
gccttctgca cctgccaccc ctgagccagc gcgggcgccc gagcgagtca tggccaacgc 240
ggggctgcag ctgttgggct tcattctcgc cttcctggga tggatcggcg ccatcgtcag 300
cactgccctg ccccagtgga ggatttactc ctatgccggc gacaacatcg tgacgcccag 360
gccatgtacg aggggctgtg gatgtcctgc gtgtcgcaga gcaccgggca gatccagtgc 420
aaagtctttg actccttgct gaatctgagc agcacattgc aagcaacccg tgccttgatg 480
gtggttggca tcctcctggg agtgatagca atctttgtgg ccaccgttgg catgaagtgt 540
atgaagtgct tggaagacga tgaggtgcag aagatgagga tggctgtcat tgggggcgcg 600
atatttcttc ttgcaggtct ggctatttta gttgccacag catggtatgg caatagaatc 660
gttcaagaat tctatgaccc tatgacccca gtcaatgcca ggtacgaatt tggtcaggct 720
ctcttcactg gctgggctgc tgcttctctc tgccttctgg gaggtgccct actttgctgt 780
tcctgtcccc gaaaaacaac ctcttaccca acaccaaggc cctatccaaa acctgcacct 840
tccagcggga aagactacgt gtgacacaga ggcaaaagga gaaaatcatg ttgaaacaaa 900
ccgaaaatgg acattgagat actatcatta acattaggac cttagaattt tgggtattgt 960
aatctgagta tggtatacaa acaacaaaca aacaaaaaac ccatgtgtta aaatactcag 1020
tgctaaacat ggcttaatct tattttatct tctttcctca atataggagg gaagattttt 1080
ccatttgtat tactgcttcc cattgagtaa tcatactcaa atgggggaag gggtgctcct 1140
taaatatata tagatatgta tatatacatg tttttctatt aaaaatagac agtaaaatac 1200
tattctcatt atgttgatac tagcatactt aaaatatctc taaaatagtt aaatgtattt 1260
aattccatat tgatgaagat gtttattggt atattttctt tttcgtcctt atatacatat 1320
gtaacagtca aatatcattt actcttcttc attagctttg ggtgcctttg ccacaagacc 1380
tagcctaatt taccaaggat gaattctttc aattcttcat gcgtgcccag caaaaaaaaa 1440
aaaaaa 1446

<210> 47
<211> 2334
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (2278)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2290)
<223> n equals a,t,g, or c

<400> 47
gggagtgtgg ctgcagaacc caggtggcag ggctttcctc aggcccctta ctcctgacct 60
ggacgaggcc ggggcttcct caaggaggct ctctgactgc caccctgcc tgcctgcccg 120
gccctgcaca acatgcagcc ctccggcctc gagggtcccg gcacgtttgg tcggtggcct 180
ctgctgagtc tgctgctcct gctgctgctg ctccagcctg taacctgtgc ctacaccacg 240
ccaggccccc ccagagccct caccacgctg ggcgccccca gagcccacac catgccgggc 300
acctacgctc cctcgaccac actcagtagt cccagcaccc agggcctgca agagcaggca 360
cgggccctga tgcgggactt cccgctcgtg gacggccaca acgacctgcc cctggtccta 420
aggcaggttt accagaaagg gctacaggat gttaacctgc gcaatttcag ctacggccag 480
accagcctgg acaggcttag agatggcctc gtgggcgccc agttctggtc agcctatgtg 540
ccatgccaga cccaggaccg ggatgccctg cgcctcaccc tggagcagat tgacctcata 600
cgccgcatgt gtgcctccta ttctgagctg gagcttgtga cctcggctaa agctctgaac 660

gacactcaga aattggcctg cctcatcggt gtagagggtg gccactcgct ggacaatagc 720
ctctccatct tacgtacctt ctacatgctg ggagtgcgct acctgacgct cacccacacc 780
tgcaacacac cctgggcaga gagctccgct aagggcgtcc actccttcta caacaacatc 840
agcgggctga ctgactttgg tgagaaggtg gtggcagaaa tgaaccgcct gggcatgatg 900
gtagacttat ccatgtctc agatgctgtg gcacggcggg ccctggaagt gtcacaggca 960
cctgtgatct tctcccactc ggctgcccgg ggtgtgtgca acagtgctcg gaatgttcct 1020
gatgacatcc tgcagcttct gaagaagaac ggtggcgtcg tgatggtgtc tttgtccatg 1080
ggagtaatac agtgcaaccc atcagccaat gtgtccactg tggcagatca cttcgaccac 1140
atcaaggctg tcattggatc caagttcatc gggattggtg gagattatga tggggccggc 1200
aaattccctc aggggctgga agacgtgtcc acatacccag tcctgataga ggagttgctg 1260
agtcgtggct ggagtgagga agagcttcag ggtgtccttc gtggaaacct gctgcgggtc 1320
ttcagacaag tggaaaaggt acaggaagaa aacaaatggc aaagcccctt ggaggacaag 1380
ttcccggatg agcagctgag cagttcctgc cactccgacc tctcacgtct gcgtcagaga 1440
cagagtctga cttcaggcca ggaactcact gagattccca tacactggac agccaagtta 1500
ccagccaagt ggtcagtctc agagtcctcc ccccacatgg ccccagtcct tgcagttgtg 1560
gccaccttcc cagtccttat tctgtggctc tgatgaccca gttagtcctg ccagatgtca 1620
ctgtagcaag ccacagacac cccacaaagt tcccctgttt gcaggcacaa atatttcctg 1680
aaataaatgt tttggacata gaaaaaaaaa aaaaaaaaag ggcggccgct ctagaggatc 1740
cctcgagggg cccaagctta cgcgtgcatg cgacgtcata gctctctccc tatagtgagt 1800
cgtattataa gctaggcact ggccgtcgtt ttacaacgtc gtgactggga gatctgctag 1860
cttgggatct ttgtgaagga accttacttc tgtggtgtga cataattgga caaactacct 1920
acagagattt aaagctctaa ggtaaatata aaatttttaa gtgtataatg tgttaaacta 1980
gctgcatatg cttgctgctt gagagttttg cttactgagt atgatttatg aaaatattat 2040
acacaggagc tagtgattct aattgtttgt gtattttaga ttcacagtcc caaggctcat 2100
ttcaggcccc tcagtcctca cagtctgttc atgatcataa tcagccatac cacatttgta 2160
gaggttttac ttgctttaaa aaacctycca cacctccccc tgaacctgaa acataaaatg 2220
aatgcaattg gtggtggtaa cttggttaat ggagcttata atggtaccaa taaagcantg 2280
catcacaaan ttcccaaata aagcattttt tcctggaatt taaatggggg ttgg 2334

<210> 48
<211> 930
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (6)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (14)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (18)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (33)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (873)

<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (930)
<223> n equals a,t,g, or c

<400> 48
gggggnaaat ttancccntc aacttaaaag ggnaacaaaa agcttggagc ttcaaccgsg 60
gktggcggcc cgttytaraa ctagggatcc cccgggctgc aaggtcaaaa ataataataa 120
atgttagcat ctcactcagt tctgtgtgat ccagctcctg ccaacccatc tgatgagctc 180
ctggttcact ctccttgctt cctgctttca cctcctgtgg cccctgtccc gttcttccca 240
tgtgccaagc tcattccagc ccccagacct gtccgctact tttctcctcc agatcttagg 300
ttaggtaaca ctccagcccc ctctgagatc acctacacac cttctgttca cttctgtcat 360
cctattgcta aattactttg tctcaaggtt cgtaatctct gtgagggagt cctctctgct 420
gcttttccca aggcctgacg cacaccgggc acccagttaa catctttaat atattgggct 480
ttcgtgctac ctgagggctt ctagaaagct ctctacatcc aagtgagagg ctggctctcc 540
aggaggtacc tgtgctctca gcacaaatgg agcccctctg ggaacaggaa cctgcttcag 600
ccctttagca tacacaggcc tgccacrtgc tggggacagg gccacctgag tcccttgcct 660
ggtcccatgt gtggtgccag ctccctcccc tgatgattgt tctctctctt tcccaggacc 720
gagatctcac tcagcagaca gcagcagccc gggagcctga gctcaggagg aactcttacc 780
tggaaattgg gaactgtatg gagactccaa actgacttct ttcaaaaaaa aaaaaaaaaa 840
aatttttta gctttgacaa acacacaaaa gtngtaataa agagagccct ccttgtcaac 900
ccaaaatgtg agccccctgt ggaaaaccan 930

<210> 49
<211> 1638
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (91)
<223> n equals a,t,g, or c

<400> 49
cccgggtcga cccacgcgtc cgcagctctc cgtggacaat gggctgtggc gtgtgaccct 60
gtgcatgctg gccttcccgc tgctcctcac nggcctcatc tccttcaggg agaagaggct 120
gcaggatgtg ggcacccccg cggcccgcgc ccgtgccttc ttcaccgcac ccgtggtggt 180
cttccacctg aacatcctct cctacttcgc cttcctctgc ctgttcgcct acgtgctcat 240
ggtggacttc cagcctgtgc cctcctggtg cgagtgtgcc atctacctct ggctcttctc 300
cttggtgtgc gaggagatgc ggcagctctt ctatgaccct gacgagtgcg ggctgatgaa 360
gaaggcagcc ttgtacttca gtgacttctg gaataagctg gacgtcggcg caatcttgct 420
cttcgtggca gggctgacct gcaggctcat cccggcgacg ctgtaccccg ggcgcgtcat 480
cctctctctg gacttcatcc tgttctgcct ccggctcatg cacattttta ccatcagtaa 540
gacgctgggg cccaagatca tcattgtgaa gcggatgatg aaggacgtct tcttcttcct 600
cttcctgctg gctgtgtggg tggtgtcctt cggggtggcc aagcaggcca tcctcatcca 660
caacgagcgc cgggtggact ggctgttccg aggggccgtc taccactcct acctcaccat 720
cttcgggcag atcccgggct acatcgacgg tgtgaacttc aacccggagc actgcagccc 780
caatggcacc gacccctaca agcctaagtg ccccgagagc gacgcgacgc agcagaggcc 840
ggccttccct gagtggctga cggtcctcct actctgcctc tacctgctct tcaccaacat 900
cctgctgctc aacctcctca tcgccatgtt caactacacc ttccagcagg tgcaggagca 960
cacggaccag atttggaagt tccagcgcca tgacctgstc gaggagtacc acggccgccc 1020
cgccgcgccg ccccccttca tcctcctcag ccacctgcag ctcttcatca agagggtggt 1080
cctgaagact ccggccaaga ggcacaagca gctcaagaac aagctggaga agaacgagga 1140
ggcggccctg ctatcctggg agatctacct gaaggagaac tacttccaga accgacagtt 1200
ccagcaaaag cagcggcccg agcagaagat cgaggacatc agcaataagg ttgacgccat 1260

ggtggacctg ctggacctgg acccactgaa gaggtcgggc tccatggagc agaggttggc 1320
ctccctggag gagcaggtgg cccagacagc ccgagccctg cactggatcg tgaggacgct 1380
gcgggccagc ggcttcagct cggaggcgga cgtccccact ctggcctccc agaaggccgc 1440
ggaggagccg gatgctgagc cgggaggcag gaagaagacg gaggagccgg gcgacagcta 1500
ccacgtgaat gcccggcacc tcctctaccc caactgccct gtcacgcgct tcccgtgccc 1560
aacgagaagg tgccctggga gacggagttc ctgatctatg acccaccctt ttacacggca 1620
gagaggaagg acgcggcc 1638

<210> 50
<211> 1275
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1275)
<223> n equals a,t,g, or c

<400> 50
ggcctggggg agagtggatc caggtgaagg gggcagaggt tacctgtgca gggatcagtc 60
catttattcc tattttctct gacagctgga tgaaagggac tgggagtttt cgtcctcttg 120
aataagaact cgacaacaga gtgggaactt tctgtcttgt gatccattgc ctggctttct 180
gtatttgtct gaatgctttc acgggagtgt gtygcactgg agcacagagg acactcgatc 240
gtgcggcgcg cagggcgggg ggccgccgct gcctccccgc gggatggctg gcactgtgct 300
cggagtcggt gcgggcgtgt tcatcttagc cctgctctgg gtggcagtgc tgctgctgtg 360
tgtgctgctg tccagagcct ccggggcggc gaggttctct gtcatttttt attcttcggt 420
gctgtgatca tcacatcagt tctgttgctt ttcccgcgag ctggtgaatt cccagcccca 480
gaagtggaag ttaagattgt ggatgacttt ttcattggcc gctatgtcct gctggctttc 540
cttagtgcca tcttccttgg aggcctcttc ttggttttaa tccattatgt tctggagccg 600
atctatgcca aaccactgca ctcctactga ccactcttca ggaawacgaa aacctgttct 660
ctccttcatt gtgatgacat tgatgagcag gaaggcacta ttcagagcct tgttttgaca 720
gccctcatgc cttaaggtta gaggagtatc tgtccatcac taagacaaat ctctggagtc 780
ctggcttcca gaaacaggat tgccaaattg tccctgtggg gctagattct taccagctta 840
agaaggatat tgctatcttc ttagtacccg taccttagga tttccaactg ttttgaaagg 900
gaaatagtaa cagtgatctg cttagagtgg attttcactc aagtccttag taagtggatt 960
ttggggaaaa aagcacmtgg gcttctggtt ctttttgata atatataaaa ttattcatta 1020
tgaggttgca gttgtttgca aaggagaggc actcaaattt gaaaggttat tttaatgtga 1080
taatttggaa gacttactca gatgttggtc attgaccact ctgtgcatat atttctgcag 1140
agctctgtga aggcaatgag tgtcacttcc ctctgctcta ataaagcaat aaataataaa 1200
aaaaaaaaaa aaaaaaaact cgaggggggg cccggtaccc aattcgccct atggtgagtc 1260
gaattggttt ccagn 1275

<210> 51
<211> 1381
<212> DNA
<213> Homo sapiens

<400> 51
gataactcag gcccggtgcc cagagcccag gaggaggcag tggccaggaa ggcacaggcc 60
tgagaagtct gcggctgagc tgggagcaaa tcccccaccc cctacctggg ggacagggtg 120
cagcggccat ggctacagca agacccccct ggatgtgggt gctctgtgct ctgatcacag 180
ccttgcttct gggggtcaca gagcatgttc tcgccaacaa tgatgtttcc tgtgaccacc 240
cctctaacac cgtgccctct gggagcaacc aggacctggg agctggggcc ggggaagacg 300
cccggtcgga tgacagcagc agccgcatca tcaatggatc cgactgcgat atgcacaccc 360
agccgtggca ggccgcgctg ttgctaaggc ccaaccagct ctactgcggg gcggtgttgg 420
tgcatccaca gtggctgctc acggccgccc actgcaggaa gaaagttttc agagtccgtc 480

tcggccacta ctccctgtca ccagtttatg aatctgggca gcagatgttc caggggtca 540
aatccatccc ccaccctggc tactcccacc ctggccactc taacgacctc atgctcatca 600
aactgaacag aagaattcgt cccactaaag atgtcagacc catcaacgtc tcctctcatt 660
gtccctctgc tgggacaaag tgcttggtgt ctggctgggg gacaaccaag agcccccaag 720
tgcacttccc taaggtcctc cagtgcttga atatcagcgt gctaagtcag aaaaggtgcg 780
aggatgctta cccgagacag atagatgaca ccatgttctg cgccggtgac aaagcaggta 840
gagactcctg ccagggtgat tctgggggggc ctgtggtctg caatggctcc ctgcagggac 900
tcgtgtcctg gggagattac ccttgtgccc ggcccaacag accgggtgtc tacacgaacc 960
tctgcaagtt caccaagtgg atccaggaaa ccatccaggc caactcctga gtcatcccag 1020
gactcagcac accggcatcc ccacctgctg cagggacagc cctgacactc ctttcagacc 1080
ctcattcctt cccagagatg ttgagaatgt tcatctctcc agcccctgac cccatgtctc 1140
ctggactcag ggtctgcttc ccccacattg ggctgaccgt gtctctctag ttgaaccctg 1200
ggaacaattt ccaaaactgt ccagggcggg ggttgcgtct caatctccct ggggcacttt 1260
catcctcaag ctcagggccc atcccttctc tgcagctctg acccaaattt agtcccagaa 1320
ataaactgag aagtggaatc ttaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1380
a 1381

<210> 52
<211> 1439
<212> DNA
<213> Homo sapiens

<400> 52
cccacgcgtc cgcaggagag gtgtctgtgc gtcctgcacc cacatctttc tctgtccccc 60
ccttgccctg tctggaggct gctagactcc tatcttctga attctatagt gcctgggtct 120
cagcgcagtg ccgatggtgg cccgtccttg tggttcctct ctacttgggg aaatcaggtg 180
cagcggccat ggctacagca agacccccct ggatgtgggt gctctgtgct ctgatcacag 240
ccttgcttct gggggtcaca gagcatgttc tcgccaacaa tgatgtttcc tgtgaccacc 300
cctctaacac cgtgccctct gggagcaacc aggacctggg agctggggcc gggggaagac 360
gcccggtcgg atgacagcag cagccgcatc atcaatggat ccgactgcga tatgcacacc 420
cagccgtggc aggccgcgct gttgctaagg cccaaccagc tctactgcgg ggcggtgttg 480
gtgcatccac agtggctgct cacggccgcc cactgcagga agaaagtttt cagagtccgt 540
ctcggccact actccctgtc accagtttat gaatctgggc agcagatgtt ccaggggtc 600
aaatccatcc cccaccctgg ctactcccac cctggccact ctaacgacct catgctcatc 660
aaactgaaca gaagaattcg tcccactaaa gatgtcagac ccatcaacgt ctcctctcat 720
tgtccctctg ctgggacaaa tgcttggtgt ctggctgggg gacaaccaag acccccaagt 780
gcacttccct aaggtcctcc agtgcttgaa tatcacgtgc taagtcagaa aaggtgcgag 840
gatgcttacc cgagacagat agatgacacc atgttctgcg ccggtgacaa agcaggtaga 900
gactcctgcc agggtgattc tgggggggcct gtggtctgca atggctccct gcagggactc 960
gtgtcctggg gagattaccc ttgtgcccgg cccaacagac cgggtgtcta cacgaacctc 1020
tgcaagttca ccaagtggat ccaggaaacc atccaggcca actcctgagt catcccagga 1080
ctcagcacac cggcatcccc acctgctgca gggacagccc tgacactcct ttcagaccct 1140
cattccttcc cagagatgtt gagaatgttc atctctccag cccctgaccc catgtctcct 1200
ggactcaggg tctgcttccc ccacattggg ctgaccgtgt ctctctagtt gaaccctggg 1260
aacaatttcc aaaactgtcc agggcggggg ttgcgtctca atctccctgg ggcactttca 1320
tcctcaagct cagggcccat cccttctctg cagctctgac ccaaatttag tcccagaaat 1380
aaactgagaa gtggaatctt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 1439

<210> 53
<211> 1329
<212> DNA
<213> Homo sapiens

<400> 53
acccacgmgc ggaggtcagt ractgccggr agtcctgcag gggcggggcg gcgccaagcg 60
cagggagccc ggctgagtgg cagcccagat tgaagatgga tacgtgacaa tcccagggac 120

cgctgcactg acttcatttc cttagacaag acacagtgta gggcccggcc cgtgttggcc 180
ccaggactcc tttggaatat agctgtggac aatgaatcct gcgagcratg ggggcacatc 240
agagagcatt tttgacctgg actatgcatc ctgggggatc cgctccacgc tgatggtcgc 300
tggctttgtc ttctacttgg gcgtctttgt ggtctgccac cagctgtcct cttccctgaa 360
tgccacttac cgttctttgg tggccagaga gaaggtcttc tgggacctgg cggccacgcg 420
tgcagtcttt ggtgttcaga gcacagccgc agctgtgggc tctgctgggg gaccctgtgc 480
tgcatgccga caaggcgcgt ggccagcaga actggtgctg gtttcacatc acgacagcaa 540
cgggattctt ttgctttgaa aatgttgcag tccacctgtc caacttgatc ttccggacat 600
ttgacttgtt tctggttatc caccatctct ttgcctttct tgggtttctt ggctgcttgg 660
tcaatctcca agctggccac tatctagcta tgaccacgtt gctcctggag atgagcacgc 720
cctttacctg cgtttcctgg atgctcttaa aggcgggctg gtccgagtct ctgttttgga 780
agctcaacca gtggctgatg attcacatgt ttcactgccg catggttcta acctaccaca 840
tgtggtgggt gtgtttctgg cactgggacg gcctggtcag cagcctgtat ctgcctcatt 900
tgacactgtt ccttgtcgga ctggctctgc ttacgctaat cattaatcca tattggaccc 960
ataagaagac tcagcagctt ctcaatccgg tggactggaa cttcgcacag ccagaagcca 1020
agagcaggcc agaaggcaac gggcagctgc tgcggaagaa gaggccatag ctgctccagc 1080
cggggctccg gggcggcagc agagctggca caccgattct gggaagcccc gcgaatgatg 1140
gcttttgaat taatgaggca gtgaatgttt tgtgtttact tctaagggaa atactaactt 1200
tctttcgcat tagtattaat tttgaagtag ctacaaagta tttttaagaa attataattt 1260
tatgactgtc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaag 1320
ggcggccgc 1329

<210> 54
<211> 1368
<212> DNA
<213> Homo sapiens

<400> 54
ctgaattgcg gccgtatgcg cggctctgtg gagtgcacct ggggttgggg gcactgtgcc 60
cccagccccc tgctcctttg gactctactt ctgtttgcag ccccatttgg cctgctgggg 120
gagaagaccc gccagctgct tgagtttgac agcaccaacg tgtccgatac ggcagcaaag 180
cctttgggaa gaccatatcc tccatactcc ttggccgatt tctcttggaa caacatcact 240
gattcattgg atcctgccac cctgagtgcc acatttcaag gccaccccat gaacgaccct 300
accaggactt ttgccaatgg cagcctggcc ttcaggtcca ggccttttcc aggtccagcc 360
gaccagccca acccccctcgc ctcctgcaca cagcagacac ctgtcagcta gaggtggccc 420
tgattggagc ctctccccgg ggaaaccgtt ccctgtttgg gctggaggta gccacattgg 480
gccagggccc tgactgcccc tcaatgcagg agcagcactc catcgacgat gaatatgcac 540
cggccgtctt ccagttggac cagctactgt ggggctccct cccatcaggc tttgcacagt 600
ggcgaccagt ggcttactcc cagaagccgg ggggccgaga atcagccctg ccctgccaag 660
cttcccctct tcatcctgcc ttagcatact ctcttcccca gtcacccatt gtccgagcct 720
tctttgggtc ccagaataac ttctgtgcct tcaatctgac gttcggggct tccacaggcc 780
ctggctattg ggaccaacac tacctcagct ggtcgatgct cctgggtttg ggcttccctc 840
cagtggacgg cttgtcccca ttagtcctgg gcatcatggc agtggcctgg gtgccccagg 900
gctcatgctg ctaggggggcg gcttggttct gctgctgcac cacaagaagt actcagagta 960
ccagtccata aattaaggcc cgctctctgg agggaaggac attactgaac ctgtcttgct 1020
gtgcctcgaa actctggagg ttggagcatc aagttccagc cggccccttc actcccccat 1080
cttgcttttc tgtggaacct cagaggccag cctcgacttc ctggagaccc ccaggtgggg 1140
cttccttcat actttgttgg gggactttgg aggcgggcag gggacagggc tattgataag 1200
gtcccccttgg tgttgccttc ttgcatctcc acacatttcc cttggatggg acttgcaggc 1260
ctaaatgaga ggcattctga ctggttggct gccctggaag gcaagaaaat agatttattt 1320
tttttcaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa 1368

<210> 55
<211> 1256
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1240)
<223> n equals a,t,g, or c

<400> 55
tcgacccacg cgtccgagca accgcagctt ctagtatcca gactccagcg ccgccccggg 60
cgcggacccc aaccccgacc cagagcttct ccagcggcgg cgcacgagca gggctccccg 120
ccttaacttc ctccgcgggg cccagccacc ttcgggagtc cgggttgccc acctgcaaac 180
tctccgcctt ctgcacctgc cacccctgag ccagcgcggg cgcccgagcg agtcatggcc 240
aacgcggggc tgcagctgtt gggcttcatt ctcgccttcc tgggatggat cggcgccatc 300
gtcagcactg ccctgcccca gtggaggatt tactcctatg ccggcgacaa catcgtgacc 360
gcccaggcca tgtacgaggg gctgtggatg tcctgcgtgt cgcagagcac cgggcagatc 420
cagtgcaaag tctttgactc cttgctgaat ctgagcagca cattgcaagc aacccgtgcc 480
ttgatggtgg ttggcatcct cctgggagtg atagcaatct ttgtggccam cgttggcatg 540
aagtgtatga agtgcttgga agacgatgag gtgcagaaga tgaggatggc tgtcattggg 600
ggcgcgatat ttcttcttgc aggtctggct attttagttg ccacagcatg gtatggcaat 660
agaatcgttc aagaattcta tgaccctatg accccagtca atgccaggta cgaatttggt 720
caggctctct tcactggctg ggctgctgct tctctctgcc ttctgggagg tgccctactt 780
tgctgttcct gtccccgaaa aacaacctct tacccaacac caaggcccta tccaaaacct 840
gcaccttcca gcgggaaaga ctacgtgtga cacagaggca aaaggagaaa atcatgttga 900
aacaaaccga aaatggacat tgagatacta tcattaacat taggacctta gaattttggg 960
tattgtaatc tgaagtatgg tattacaaaa caaacaaaca aacaaaaaac ccatgtgtta 1020
aaatactcag tgctaaacat ggcttaatct tattttatct tctttcctca atataggagg 1080
gaagattttt ccatttgtat tactgcttcc cattgagtaa tcatactcaa ctgggggaag 1140
gggtgctcct taaatatata tagatatgta tatatacatg tttttctatt aaaaatagac 1200
agtaaaatwc taaaaaaaaa aaaaaaamcy cggggggggn ccggtaccca ttcgcc 1256

<210> 56
<211> 143
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (143)
<223> Xaa equals stop translation

<400> 56
Met Ser Gly Ile Ser Gly Cys Pro Phe Phe Leu Trp Gly Leu Leu Ala
1 5 10 15

Leu Leu Gly Leu Ala Leu Val Ile Ser Leu Ile Phe Asn Ile Ser His
20 25 30

Tyr Val Glu Lys Gln Arg Gln Asp Lys Met Tyr Ser Tyr Ser Ser Asp
35 40 45

His Thr Arg Val Asp Glu Tyr Tyr Ile Glu Asp Thr Pro Ile Tyr Gly
50 55 60

Asn Leu Asp Asp Met Ile Ser Glu Pro Met Asp Glu Asn Cys Tyr Glu
65 70 75 80

Gln Met Lys Ala Arg Pro Glu Lys Ser Val Asn Lys Met Gln Glu Ala
85 90 95

Thr Pro Ser Ala Gln Ala Thr Asn Glu Thr Gln Met Cys Tyr Ala Ser
100 105 110

Leu Asp His Ser Val Lys Gly Lys Arg Arg Ser Pro Gly Asn Arg Ile
115 120 125

Leu Ile Ser Gln Thr Arg Met Glu Met Ser Asn Tyr Met Gln Xaa
130 135 140

<210> 57
<211> 51
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (51)
<223> Xaa equals stop translation

<400> 57
Met Ala Leu Met Trp Ser Leu Trp Tyr Phe Asn Ser Val Phe Ile Ile
1 5 10 15

Ser Cys Val Ser Gly Lys Ile Val Leu Thr Phe Pro Leu Tyr Thr Thr
20 25 30

Val Cys Ser Tyr Gly Ala Leu Asn Cys Leu Thr Glu Glu Pro Ser Ser
35 40 45

Val Phe Xaa
50

<210> 58
<211> 102
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (102)
<223> Xaa equals stop translation

<400> 58
Met Glu Glu Ala Ile Leu Val Pro Cys Val Leu Gly Leu Leu Leu Leu
1 5 10 15

Pro Ile Leu Ala Met Leu Met Ala Leu Cys Val His Cys His Arg Leu
20 25 30

Pro Gly Ser Tyr Asp Ser Thr Ser Ser Asp Ser Leu Tyr Pro Lys Gly
35 40 45

His Pro Val Gln Thr Ala Ser His Gly Cys Pro Leu Ala Thr Cys Leu
50 55 60

Pro Thr Cys His Leu Leu Pro Thr Pro Glu Pro Ala Arg Pro Ala Pro
65 70 75 80

His Pro Lys Ile Pro Ala Ala Pro Trp Gly Leu Pro Pro Asp Ala Ile
85 90 95

Phe Pro Ala Gly Phe Xaa
100

<210> 59
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (31)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation

<400> 59
Met Ser Cys Ile Gly Arg Met Arg Leu Ile Cys Phe Ile Ile Leu Arg
1 5 10 15

Ile Cys Gly Leu Glu His Leu Phe Gly Asn Met Gly Leu Gly Xaa Lys
20 25 30

Asn Gly His Leu Pro Gly His Tyr Gly His Ser Leu Glu Phe Phe Xaa
35 40 45

<210> 60
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (98)
<223> Xaa equals stop translation

<400> 60
Met Ile Leu Leu Leu Ser Leu Phe Gln Gly Val Arg Gly Ser Leu Gly
1 5 10 15

Ser Pro Gly Asn Arg Glu Asn Lys Glu Lys Lys Val Phe Ile Ser Leu
20 25 30

Val Gly Ser Arg Gly Leu Gly Cys Ser Ile Ser Ser Gly Pro Ile Gln
35 40 45

Lys Pro Gly Ile Phe Ile Ser His Val Lys Pro Gly Ser Leu Ser Ala
50 55 60

Glu Val Gly Leu Glu Ile Gly Asp Gln Ile Val Glu Val Asn Gly Val
65 70 75 80

Asp Phe Ser Asn Leu Asp His Lys Glu Leu Gln Leu Ala Gly Ser Cys
85 90 95

Ser Xaa

<210> 61
<211> 52
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (52)
<223> Xaa equals stop translation

<400> 61
Met Trp Phe Arg Cys Phe Leu Leu Ile Phe Val Ser Ser Val Thr Leu
1 5 10 15

Thr Gly Asp Phe Arg Asn Met Lys Lys Pro Ser Ser Leu Cys Leu Phe
20 25 30

Arg Gln Gly Leu Met Ser Ala Ser Glu Val Ser Gly Ser Gly Ser Gly
35 40 45

Glu Gly Asp Xaa
50

<210> 62
<211> 52
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (52)
<223> Xaa equals stop translation

<400> 62
Met Tyr Cys Leu Cys Gly Leu Leu Leu Gln Ala Leu Leu Arg Leu Cys
1 5 10 15

Asn Gly Tyr Lys Thr Gln Lys Asn His Arg Glu Leu Arg Met Cys Gly
20 25 30

Ile Ile Ala Gln Gly Lys Ser Arg Trp Gln Leu His Cys Tyr Pro Gly
35 40 45

Met Lys Ser Xaa
50

<210> 63
<211> 41
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (41)
<223> Xaa equals stop translation

<400> 63
Met Phe Val Phe Leu Ser Val Leu Tyr Ser Leu Ser Leu Glu Tyr Met
1 5 10 15

Phe Leu Phe Val Phe Gly Lys Lys Ile Ser Phe Thr Ser Leu His Ser
20 25 30

Asp Gln Leu Gly Lys Lys Lys Ala Xaa
35 40

<210> 64
<211> 42
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (42)
<223> Xaa equals stop translation

<400> 64
Met Tyr Met Lys Gln Val Val Ala Cys Arg Asp Gln Leu Ile Leu Val
1 5 10 15

Leu Trp Leu Ile Glu Leu Leu Cys Ile Gln Gly Phe Cys Lys Ser Lys
20 25 30

Ser Asp Phe Ser Ser Arg Ile Tyr Ser Xaa
35 40

<210> 65
<211> 183
<212> PRT
<213> Homo sapiens

<400> 65
Met Ser Lys Glu Pro Leu Ile Leu Trp Leu Met Ile Glu Phe Trp Trp
1 5 10 15

Leu Tyr Leu Thr Pro Val Thr Ser Glu Thr Val Val Thr Glu Val Leu
20 25 30

```
Gly His Arg Val Thr Leu Pro Cys Leu Tyr Ser Ser Trp Ser His Asn
         35                  40                  45

Ser Asn Ser Met Cys Trp Gly Lys Asp Gln Cys Pro Tyr Ser Gly Cys
     50                  55                  60

Lys Glu Ala Leu Ile Arg Thr Asp Gly Met Arg Val Thr Ser Arg Lys
 65                  70                  75                  80

Ser Ala Lys Tyr Arg Leu Gln Gly Thr Ile Pro Arg Gly Asp Val Ser
                 85                  90                  95

Leu Thr Ile Leu Asn Pro Ser Glu Ser Asp Ser Gly Val Tyr Cys Cys
            100                 105                 110

Arg Ile Glu Val Pro Gly Trp Phe Asn Asp Val Lys Ile Asn Val Arg
        115                 120                 125

Leu Asn Leu Gln Arg Ala Ser Thr Thr Thr His Arg Thr Ala Thr Thr
    130                 135                 140

Thr Thr Arg Arg Thr Thr Thr Thr Ser Pro Thr Thr Thr Arg Gln Met
145                 150                 155                 160

Thr Thr Thr Pro Ala Ala Leu Pro Thr Thr Lys Lys Lys Lys Lys Lys
                165                 170                 175

Lys Lys Lys Lys Lys Lys Lys
            180
```

<210> 66
<211> 58
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (58)
<223> Xaa equals stop translation

<400> 66

```
Met Leu Tyr Phe Cys Ser Ser Ile Trp Phe Gly Ile Tyr Phe Val Ala
  1               5                  10                  15

Leu Ile Thr Val Phe Leu Lys Thr Leu Pro Pro Leu Thr Val Gly Lys
             20                  25                  30

Gly Pro Phe Ser Gly Lys Phe Val Ala Phe Phe Phe Phe Leu Lys Glu
         35                  40                  45

Ser Cys Ser Leu Leu Ser Ile Val Phe Xaa
     50                  55
```

<210> 67
<211> 100
<212> PRT

<213> Homo sapiens

<400> 67

```
Met Gln Phe Cys Glu Leu Trp Val Pro Leu Leu Ser Thr Leu Leu Asn
  1               5                  10                  15

Thr Trp Gln Asn Leu Thr Leu Gly Cys Pro Ser Pro Asp Ser Lys Ser
             20                  25                  30

Lys Ser Ser Pro Asp Pro Arg Ala Cys Pro Leu Phe Pro Ser Phe Leu
         35                  40                  45

Ser Phe Phe Leu Val Ser Ser Phe Phe Phe Phe Phe Ser Phe Phe Phe
     50                  55                  60

Leu Ser Leu Ser Phe Phe Leu Pro Phe Phe Phe Leu Phe Ser Phe Phe
 65                  70                  75                  80

Leu Ser Leu Ser Leu Ser Phe Phe Gln Asp Pro Val Gln Lys Lys Lys
                 85                  90                  95

Lys Lys Thr Arg
            100
```

<210> 68
<211> 74
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (74)
<223> Xaa equals stop translation

<400> 68

```
Met Phe Tyr Leu Tyr Ser Ile Phe Gln Val Leu Val Trp Leu Cys Gln
  1               5                  10                  15

Ala Lys His Leu Ser Gln Ile Ser Ala Arg Ser Ser Arg Arg Leu Trp
             20                  25                  30

Arg Leu Ser Leu Ile Thr Phe Pro Pro Tyr Leu Ala Thr Ser Leu Ser
         35                  40                  45

His Gly Pro His Val Cys Leu Gln Thr Leu Gly Tyr Glu Ser Cys Glu
     50                  55                  60

His Thr Asp Leu Cys Phe Leu His Asp Xaa
 65                  70
```

<210> 69
<211> 49
<212> PRT
<213> Homo sapiens

<220>

<221> SITE
<222> (49)
<223> Xaa equals stop translation

<400> 69
Met Gln Pro Cys Leu Phe Met Phe Val Leu Met Gly Ile Met Trp Ala
1 5 10 15

Thr Gly Ile Leu Pro Lys Ile Met Pro Ser Arg Lys Arg Cys Leu Ser
20 25 30

Ile Asp Ile Pro Ala Ala Pro Gln Ala Gly Met Cys Leu Leu Ile Leu
35 40 45

Xaa

<210> 70
<211> 46
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (46)
<223> Xaa equals stop translation

<400> 70
Met Arg Thr Leu Ala Leu Leu Val Leu Leu Phe Cys Ser Cys Thr His
1 5 10 15

Ser Ser Met Gly Trp Gly Arg Gln Ala Trp Gly Val Ala Leu Gly Glu
20 25 30

Val Arg Ser Pro Pro Ala Gln Asp Thr Val Ala Lys Thr Xaa
35 40 45

<210> 71
<211> 64
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (64)
<223> Xaa equals stop translation

<400> 71
Met Cys Ala Trp His Cys Val His Leu Ala Leu Cys Val Val Gly Met
1 5 10 15

Leu Phe Leu Leu Ser Val Thr Ser Ser Gln Phe Cys Lys Gln Arg Gln
20 25 30

Asn His Ala Leu Pro Leu Lys Pro Ile Gly Phe Lys Cys His Leu Phe
35 40 45

Asp Asp Ala Phe Pro Ile Thr Pro Phe Asp Thr Ser His Gly Thr Xaa
50 55 60

<210> 72
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation

<400> 72
Met Phe Met Tyr Val Trp Cys Pro Leu Val Leu Phe Phe Phe Leu Leu
1 5 10 15

Val Phe Glu Leu Val Leu Asn Arg Ile Leu Ser Gly Phe Leu Lys Tyr
20 25 30

Phe His Phe His His Gly Tyr Asn Lys Phe Ala Ala Cys Pro Asn Xaa
35 40 45

<210> 73
<211> 49
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (49)
<223> Xaa equals stop translation

<400> 73
Met Val Ser Pro Trp Leu Pro Leu Leu Val Ser Leu Phe His Leu Leu
1 5 10 15

Asn Cys Leu Arg Gly Val Gly Thr Ser Gly Gln Ser Leu Gly Leu Pro
20 25 30

Ser Ser Ser Phe Pro Pro Thr Pro Glu His Lys Ala Thr Ala Arg Asp
35 40 45

Xaa

<210> 74
<211> 47

<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (47)
<223> Xaa equals stop translation

<400> 74
Gly Lys Thr Leu Tyr Leu Pro Val Cys Leu Ser Phe Leu His Ser Pro
1 5 10 15

Ala Ser Thr Phe Leu Pro Trp Asn Gln Gly Phe Leu Ser Pro Phe Ala
20 25 30

Phe Ser Thr Leu Gly Thr Pro Gly Ala Lys Gln Phe Ser Ile Xaa
35 40 45

<210> 75
<211> 59
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (59)
<223> Xaa equals stop translation

<400> 75
Met Val Ser Leu Cys Ser Gly Leu Pro Ser Ser Cys Leu Leu Leu Gly
1 5 10 15

Ser Thr Ala Ala Ile Ile Gln Arg Gln Val Cys Leu Phe Gln Gly Ala
20 25 30

Arg Gln Trp Asn Pro Val Ser Glu Phe Leu Arg Ala His His His Cys
35 40 45

Gly Asn Arg Ala Gly Leu Pro Ala Val Leu Xaa
50 55

<210> 76
<211> 90
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (90)
<223> Xaa equals stop translation

<400> 76
Met Ala Lys Arg Thr Phe Ser Asn Leu Glu Thr Phe Leu Ile Phe Leu
1 5 10 15

Leu Val Met Met Ser Ala Ile Thr Val Ala Leu Leu Ser Leu Leu Phe

20 25 30

Ile Thr Ser Gly Thr Ile Glu Asn His Lys Asp Leu Gly Gly His Phe
35 40 45

Phe Ser Thr Thr Gln Ser Pro Pro Ala Thr Gln Gly Ser Thr Ala Ala
50 55 60

Gln Arg Ser Thr Ala Thr Gln His Ser Thr Ala Thr Gln Ser Ser Asn
65 70 75 80

Ser Gln Leu Lys Leu Leu Gln Cys Leu Xaa
85 90

<210> 77
<211> 44
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (44)
<223> Xaa equals stop translation

<400> 77
Met Ser Ser Trp Phe Thr Leu Leu Ala Ser Cys Phe His Leu Leu Trp
1 5 10 15

Pro Leu Ser Arg Ser Ser His Val Pro Ser Ser Phe Gln Pro Pro Asp
20 25 30

Leu Ser Ala Thr Phe Leu Leu Gln Ile Leu Gly Xaa
35 40

<210> 78
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation

<400> 78
Met Leu Ile Ser Val Asp Ser Asn Val Pro Val Val Phe Leu Leu Leu
1 5 10 15

Phe Ile Leu Val Ile Leu Cys His Met Glu Cys Lys Gly His Ile Tyr
20 25 30

Ile Cys Val Cys Val Cys Val Tyr Met Tyr Ile Phe Lys Asn Ile Xaa
35 40 45

<210> 79
<211> 60
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (60)
<223> Xaa equals stop translation

<400> 79
Met Met Lys Asp Val Phe Phe Phe Leu Phe Leu Leu Ala Val Trp Val
1 5 10 15

Val Ser Phe Gly Val Ala Lys Gln Ala Ile Leu Ile His Asn Glu Arg
20 25 30

Arg Val Asp Trp Leu Phe Arg Gly Pro Ser Thr Thr Pro Thr Ser Pro
35 40 45

Ser Ser Gly Arg Ser Arg Ala Thr Ser Thr Val Xaa
50 55 60

<210> 80
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation

<400> 80
Met Ala Gly Thr Val Leu Gly Val Gly Ala Gly Val Phe Ile Leu Ala
1 5 10 15

Leu Leu Trp Val Ala Val Leu Leu Leu Cys Val Leu Leu Ser Arg Ala
20 25 30

Ser Gly Ala Ala Arg Phe Ser Val Ile Phe Tyr Ser Ser Val Leu Xaa
35 40 45

<210> 81
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (48)

<223> Xaa equals stop translation

<400> 81

```
Met Ser Leu Leu Leu Pro Pro Leu Ala Leu Leu Leu Leu Leu Ala Ala
  1               5                  10                  15

Leu Val Ala Pro Ala Thr Ala Ala Thr Ala Tyr Arg Pro Asp Trp Asn
             20                  25                  30

Arg Leu Ser Gly Leu Thr Arg Ala Arg Val Glu Thr Cys Gly Gly Xaa
         35                  40                  45
```

<210> 82
<211> 293
<212> PRT
<213> *Homo sapiens*

<400> 82

```
Met Ala Thr Ala Arg Pro Pro Trp Met Trp Val Leu Cys Ala Leu Ile
  1               5                  10                  15

Thr Ala Leu Leu Leu Gly Val Thr Glu His Val Leu Ala Asn Asn Asp
             20                  25                  30

Val Ser Cys Asp His Pro Ser Asn Thr Val Pro Ser Gly Ser Asn Gln
         35                  40                  45

Asp Leu Gly Ala Gly Ala Gly Glu Asp Ala Arg Ser Asp Asp Ser Ser
     50                  55                  60

Ser Arg Ile Ile Asn Gly Ser Asp Cys Asp Met His Thr Gln Pro Trp
 65                  70                  75                  80

Gln Ala Ala Leu Leu Leu Arg Pro Asn Gln Leu Tyr Cys Gly Ala Val
                 85                  90                  95

Leu Val His Pro Gln Trp Leu Leu Thr Ala Ala His Cys Arg Lys Lys
            100                 105                 110

Val Phe Arg Val Arg Leu Gly His Tyr Ser Leu Ser Pro Val Tyr Glu
        115                 120                 125

Ser Gly Gln Gln Met Phe Gln Gly Val Lys Ser Ile Pro His Pro Gly
    130                 135                 140

Tyr Ser His Pro Gly His Ser Asn Asp Leu Met Leu Ile Lys Leu Asn
145                 150                 155                 160

Arg Arg Ile Arg Pro Thr Lys Asp Val Arg Pro Ile Asn Val Ser Ser
                165                 170                 175

His Cys Pro Ser Ala Gly Thr Lys Cys Leu Val Ser Gly Trp Gly Thr
            180                 185                 190
```

Thr Lys Ser Pro Gln Val His Phe Pro Lys Val Leu Gln Cys Leu Asn
195 200 205

Ile Ser Val Leu Ser Gln Lys Arg Cys Glu Asp Ala Tyr Pro Arg Gln
210 215 220

Ile Asp Asp Thr Met Phe Cys Ala Gly Asp Lys Ala Gly Arg Asp Ser
225 230 235 240

Cys Gln Gly Asp Ser Gly Gly Pro Val Val Cys Asn Gly Ser Leu Gln
245 250 255

Gly Leu Val Ser Trp Gly Asp Tyr Pro Cys Ala Arg Pro Asn Arg Pro
260 265 270

Gly Val Tyr Thr Asn Leu Cys Lys Phe Thr Lys Trp Ile Gln Glu Thr
275 280 285

Ile Gln Ala Asn Ser
290

<210> 83
<211> 89
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (89)
<223> Xaa equals stop translation

<400> 83
Met Val Ala Gly Phe Val Phe Tyr Leu Gly Val Phe Val Val Cys His
1 5 10 15

Gln Leu Ser Ser Ser Leu Asn Ala Thr Tyr Arg Ser Leu Val Ala Arg
20 25 30

Glu Lys Val Phe Trp Asp Leu Ala Ala Thr Arg Ala Val Phe Gly Val
35 40 45

Gln Ser Thr Ala Ala Ala Val Gly Ser Ala Gly Gly Pro Cys Ala Ala
50 55 60

Cys Arg Gln Gly Ala Trp Pro Ala Glu Leu Val Leu Val Ser His His
65 70 75 80

Asp Ser Asn Gly Ile Leu Leu Leu Xaa
85

<210> 84
<211> 250
<212> PRT
<213> Homo sapiens

<220>

<221> SITE
<222> (161)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (212)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (213)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (215)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (216)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (218)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (221)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (250)
<223> Xaa equals stop translation

<400> 84

```
Met Trp Arg Cys Pro Leu Gly Leu Leu Leu Leu Pro Leu Ala Gly
  1               5                  10                  15

His Leu Ala Leu Gly Ala Gln Gln Gly Arg Gly Arg Arg Glu Leu Ala
             20                  25                  30

Pro Gly Leu His Leu Arg Gly Ile Arg Asp Ala Gly Gly Arg Tyr Cys
         35                  40                  45

Gln Glu Gln Asp Leu Cys Cys Arg Gly Arg Ala Asp Asp Cys Ala Leu
     50                  55                  60

Pro Tyr Leu Gly Ala Ile Cys Tyr Cys Asp Leu Phe Cys Asn Arg Thr
 65                  70                  75                  80

Val Ser Asp Cys Cys Pro Asp Phe Trp Asp Phe Cys Leu Gly Val Pro
                 85                  90                  95
```

```
Pro Pro Phe Pro Pro Ile Gln Gly Cys Met His Gly Gly Arg Ile Tyr
            100                 105                 110

Pro Val Leu Gly Thr Tyr Trp Asp Asn Cys Asn Arg Cys Thr Cys Gln
        115                 120                 125

Glu Asn Arg Gln Trp Gln Cys Asp Gln Glu Pro Cys Leu Val Asp Pro
    130                 135                 140

Asp Met Ile Lys Ala Ile Asn Gln Gly Asn Tyr Gly Trp Gln Ala Gly
145                 150                 155                 160

Xaa His Ser Ala Phe Trp Gly Met Thr Leu Asp Glu Gly Ile Arg Tyr
                165                 170                 175

Arg Leu Gly Thr Ile Arg Pro Ser Ser Ser Val Met Asn Met His Glu
            180                 185                 190

Ile Tyr Thr Val Leu Asn Pro Gly Glu Val Leu Pro Thr Ala Phe Glu
        195                 200                 205

Ala Ser Glu Xaa Xaa Pro Xaa Xaa Phe Xaa Ser Leu Xaa Thr Lys Ala
    210                 215                 220

Thr Val Gln Ala Pro Gly Pro Ser Pro Gln Gln Leu Trp His Pro Ile
225                 230                 235                 240

Val Ser Gln Ser Ile Leu Trp Asp Thr Xaa
                245                 250
```

<210> 85
<211> 58
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (58)
<223> Xaa equals stop translation

<400> 85

```
Met Tyr Thr Lys Leu Met Leu Asn Lys Val Leu Leu Phe Trp Gln Ile
  1               5                  10                  15

Val Lys Cys Lys Val Leu Val Asp Gln Tyr Cys Tyr Asn Phe Gly Ala
             20                  25                  30

Lys Leu Leu His Ala Asp Trp Leu Trp Asp Leu Val His Phe Leu Arg
         35                  40                  45

Thr Asn Val Glu Phe Glu Lys Thr Pro Xaa
     50                  55
```

<210> 86
<211> 49
<212> PRT

<213> Homo sapiens

<220>
<221> SITE
<222> (49)
<223> Xaa equals stop translation

<400> 86
Met Phe Leu Gly Ser Leu Cys Phe Ser Leu Leu Cys His Ala Gly Cys
1 5 10 15

Gln Gly Ser Gly Gly Lys Pro Ala Val Thr Gly Leu Thr Gln Leu Pro
20 25 30

His Asn Pro Lys Gly Trp Phe His Ser His His Ala Pro Arg Pro Gln
35 40 45

Xaa

<210> 87
<211> 172
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (170)
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 87
Met Arg Gly Ser Val Glu Cys Thr Trp Gly Trp Gly His Cys Ala Pro
1 5 10 15

Ser Pro Leu Leu Leu Trp Thr Leu Leu Leu Phe Ala Ala Pro Phe Gly
20 25 30

Leu Leu Gly Glu Lys Thr Arg Gln Leu Leu Glu Phe Asp Ser Thr Asn
35 40 45

Val Ser Asp Thr Ala Ala Lys Pro Leu Gly Arg Pro Tyr Pro Pro Tyr
50 55 60

Ser Leu Ala Asp Phe Ser Trp Asn Asn Ile Thr Asp Ser Leu Asp Pro
65 70 75 80

Ala Thr Leu Ser Ala Thr Phe Gln Gly His Pro Met Asn Asp Pro Thr
85 90 95

Arg Thr Phe Ala Asn Gly Ser Leu Ala Phe Arg Val Gln Ala Phe Ser
100 105 110

Arg Ser Ser Arg Pro Ala Gln Pro Pro Arg Leu Leu His Thr Ala Asp
115 120 125

Thr Cys Gln Leu Glu Val Ala Leu Ile Gly Ala Ser Pro Arg Gly Asn
130 135 140

Arg Ser Leu Phe Gly Leu Glu Val Ala Thr Leu Gly Gln Gly Pro Asp
145 150 155 160

Cys Pro Ser Met Gln Glu Gln His Ser Xaa Glu Arg
165 170

<210> 88
<211> 174
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (174)
<223> Xaa equals stop translation

<400> 88
Met Val Phe Leu Lys Phe Phe Cys Met Ser Phe Phe Cys His Leu Cys
1 5 10 15

Gln Gly Tyr Phe Asp Gly Pro Leu Tyr Pro Glu Met Ser Asn Gly Thr
20 25 30

Leu His His Tyr Phe Val Pro Asp Gly Asp Tyr Glu Glu Asn Asp Asp
35 40 45

Pro Glu Lys Cys Gln Leu Leu Phe Arg Val Ser Asp His Arg Arg Cys
50 55 60

Ser Gln Gly Glu Gly Ser Gln Val Gly Ser Leu Leu Ser Leu Thr Leu
65 70 75 80

Arg Glu Glu Phe Thr Val Leu Gly His Gln Val Glu Gly Cys Trp Ala
85 90 95

Arg Ala Gly Gly His Gln Gln Lys His Leu Leu Arg Pro Arg Arg Gly
100 105 110

Arg Glu Leu Trp Gln Val Pro Ala Ala Gly Val Pro Pro Asp Arg Gly
115 120 125

Met Pro Thr Pro Thr Arg Thr Asn Pro Ser Leu Ser Trp Arg Ala Ser
130 135 140

Ser Ser Arg Ala Arg Asn Arg Thr Ala Gly Arg Arg Ala Gly Ser Thr
145 150 155 160

Arg Thr Phe Trp Glu Cys Trp Ser Thr Pro Gly Pro Cys Xaa
165 170

<210> 89
<211> 275
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (275)
<223> Xaa equals stop translation

<400> 89

```
Met Phe Tyr Ile Ile Gly Gly Val Ala Thr Leu Leu Leu Ile Leu Val
 1               5                  10                  15

Ile Ile Val Phe Lys Glu Lys Pro Lys Tyr Pro Pro Ser Arg Ala Gln
            20                  25                  30

Ser Leu Ser Tyr Ala Leu Thr Ser Pro Asp Ala Ser Tyr Leu Gly Ser
        35                  40                  45

Ile Ala Arg Leu Phe Lys Asn Leu Asn Phe Val Leu Leu Val Ile Thr
    50                  55                  60

Tyr Gly Leu Asn Ala Gly Ala Phe Tyr Ala Leu Ser Thr Leu Leu Asn
65                  70                  75                  80

Arg Met Val Ile Trp His Tyr Pro Gly Glu Glu Val Asn Ala Gly Arg
                85                  90                  95

Ile Gly Leu Thr Ile Val Ile Ala Gly Met Leu Gly Ala Val Ile Ser
            100                 105                 110

Gly Ile Trp Leu Asp Arg Ser Lys Thr Tyr Lys Glu Thr Thr Leu Val
        115                 120                 125

Val Tyr Ile Met Thr Leu Val Gly Met Val Val Tyr Thr Phe Thr Leu
    130                 135                 140

Asn Leu Gly His Leu Trp Val Val Phe Ile Thr Ala Gly Thr Met Gly
145                 150                 155                 160

Phe Phe Met Thr Gly Tyr Leu Pro Leu Gly Phe Glu Phe Ala Val Glu
                165                 170                 175

Leu Thr Tyr Pro Glu Ser Glu Gly Ile Ser Ser Gly Leu Leu Asn Ile
            180                 185                 190

Ser Ala Gln Val Phe Gly Ile Ile Phe Thr Ile Ser Gln Gly Gln Ile
        195                 200                 205

Ile Asp Asn Tyr Gly Thr Lys Pro Gly Asn Ile Phe Leu Cys Val Phe
    210                 215                 220

Leu Thr Leu Gly Ala Ala Leu Thr Ala Phe Ile Lys Ala Asp Leu Arg
225                 230                 235                 240

Arg Gln Lys Ala Asn Lys Glu Thr Leu Glu Asn Lys Leu Gln Glu Glu
                245                 250                 255

Glu Glu Glu Ser Asn Thr Ser Lys Val Pro Thr Ala Val Ser Glu Asp
            260                 265                 270

His Leu Xaa
```

275

<210> 90
<211> 83
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (83)
<223> Xaa equals stop translation

<400> 90
Met Lys Lys Val Ala Arg Leu Ser Ser Leu Gly His Val Val Trp Arg
1 5 10 15

Leu Tyr Ala Arg Val Leu Ala Leu Ile Thr Cys Ile Phe Trp Val Leu
20 25 30

Ala Leu Ile Ile Cys Ile Phe Thr Pro Gln Ile Phe Phe Lys His Leu
35 40 45

Leu His Ala Arg Pro Cys Ser Arg Tyr Arg Arg Tyr Asn Ser Lys Asn
50 55 60

Thr Asp Leu Ala Leu Met Lys Leu Lys Leu Leu Arg Gln Ala Asp Ser
65 70 75 80

Asp Lys Xaa

<210> 91
<211> 71
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (71)
<223> Xaa equals stop translation

<400> 91
Met Ala Asn Ala Gly Leu Gln Leu Leu Gly Phe Ile Leu Ala Phe Leu
1 5 10 15

Gly Trp Ile Gly Ala Ile Val Ser Thr Ala Leu Pro Gln Trp Arg Ile
20 25 30

Tyr Ser Tyr Ala Gly Asp Asn Ile Val Thr Pro Arg Pro Cys Thr Arg
35 40 45

Gly Cys Gly Cys Pro Ala Cys Arg Arg Ala Pro Gly Arg Ser Ser Ala
50 55 60

Lys Ser Leu Thr Pro Cys Xaa
65 70

<210> 92
<211> 486
<212> PRT
<213> Homo sapiens

<400> 92

```
Met Gln Pro Ser Gly Leu Glu Gly Pro Gly Thr Phe Gly Arg Trp Pro
  1               5                  10                  15

Leu Leu Ser Leu Leu Leu Leu Leu Leu Leu Leu Gln Pro Val Thr Cys
             20                  25                  30

Ala Tyr Thr Thr Pro Gly Pro Pro Arg Ala Leu Thr Thr Leu Gly Ala
         35                  40                  45

Pro Arg Ala His Thr Met Pro Gly Thr Tyr Ala Pro Ser Thr Thr Leu
     50                  55                  60

Ser Ser Pro Ser Thr Gln Gly Leu Gln Glu Gln Ala Arg Ala Leu Met
 65                  70                  75                  80

Arg Asp Phe Pro Leu Val Asp Gly His Asn Asp Leu Pro Leu Val Leu
                 85                  90                  95

Arg Gln Val Tyr Gln Lys Gly Leu Gln Asp Val Asn Leu Arg Asn Phe
            100                 105                 110

Ser Tyr Gly Gln Thr Ser Leu Asp Arg Leu Arg Asp Gly Leu Val Gly
        115                 120                 125

Ala Gln Phe Trp Ser Ala Tyr Val Pro Cys Gln Thr Gln Asp Arg Asp
    130                 135                 140

Ala Leu Arg Leu Thr Leu Glu Gln Ile Asp Leu Ile Arg Arg Met Cys
145                 150                 155                 160

Ala Ser Tyr Ser Glu Leu Glu Leu Val Thr Ser Ala Lys Ala Leu Asn
                165                 170                 175

Asp Thr Gln Lys Leu Ala Cys Leu Ile Gly Val Glu Gly Gly His Ser
            180                 185                 190

Leu Asp Asn Ser Leu Ser Ile Leu Arg Thr Phe Tyr Met Leu Gly Val
        195                 200                 205

Arg Tyr Leu Thr Leu Thr His Thr Cys Asn Thr Pro Trp Ala Glu Ser
    210                 215                 220

Ser Ala Lys Gly Val His Ser Phe Tyr Asn Asn Ile Ser Gly Leu Thr
225                 230                 235                 240

Asp Phe Gly Glu Lys Val Val Ala Glu Met Asn Arg Leu Gly Met Met
                245                 250                 255

Val Asp Leu Ser His Val Ser Asp Ala Val Ala Arg Arg Ala Leu Glu
            260                 265                 270
```

```
Val Ser Gln Ala Pro Val Ile Phe Ser His Ser Ala Ala Arg Gly Val
        275                 280                 285

Cys Asn Ser Ala Arg Asn Val Pro Asp Asp Ile Leu Gln Leu Leu Lys
    290                 295                 300

Lys Asn Gly Gly Val Val Met Val Ser Leu Ser Met Gly Val Ile Gln
305                 310                 315                 320

Cys Asn Pro Ser Ala Asn Val Ser Thr Val Ala Asp His Phe Asp His
                325                 330                 335

Ile Lys Ala Val Ile Gly Ser Lys Phe Ile Gly Ile Gly Gly Asp Tyr
            340                 345                 350

Asp Gly Ala Gly Lys Phe Pro Gln Gly Leu Glu Asp Val Ser Thr Tyr
        355                 360                 365

Pro Val Leu Ile Glu Glu Leu Leu Ser Arg Gly Trp Ser Glu Glu Glu
    370                 375                 380

Leu Gln Gly Val Leu Arg Gly Asn Leu Leu Arg Val Phe Arg Gln Val
385                 390                 395                 400

Glu Lys Val Gln Glu Glu Asn Lys Trp Gln Ser Pro Leu Glu Asp Lys
                405                 410                 415

Phe Pro Asp Glu Gln Leu Ser Ser Ser Cys His Ser Asp Leu Ser Arg
            420                 425                 430

Leu Arg Gln Arg Gln Ser Leu Thr Ser Gly Gln Glu Leu Thr Glu Ile
        435                 440                 445

Pro Ile His Trp Thr Ala Lys Leu Pro Ala Lys Trp Ser Val Ser Glu
    450                 455                 460

Ser Ser Pro His Met Ala Pro Val Leu Ala Val Val Ala Thr Phe Pro
465                 470                 475                 480

Val Leu Ile Leu Trp Leu
                485
```

<210> 93
<211> 44
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (44)
<223> Xaa equals stop translation

<400> 93

```
Met Ser Ser Trp Phe Thr Leu Leu Ala Ser Cys Phe His Leu Leu Trp
  1               5                  10                  15
```

```
Pro Leu Ser Arg Ser Ser His Val Pro Ser Ser Phe Gln Pro Pro Asp
            20                  25                  30

Leu Ser Ala Thr Phe Leu Leu Gln Ile Leu Gly Xaa
        35                  40
```

<210> 94
<211> 509
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (312)
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 94

```
Met Leu Ala Phe Pro Leu Leu Leu Thr Gly Leu Ile Ser Phe Arg Glu
  1               5                  10                  15

Lys Arg Leu Gln Asp Val Gly Thr Pro Ala Ala Arg Ala Arg Ala Phe
            20                  25                  30

Phe Thr Ala Pro Val Val Val Phe His Leu Asn Ile Leu Ser Tyr Phe
        35                  40                  45

Ala Phe Leu Cys Leu Phe Ala Tyr Val Leu Met Val Asp Phe Gln Pro
    50                  55                  60

Val Pro Ser Trp Cys Glu Cys Ala Ile Tyr Leu Trp Leu Phe Ser Leu
65                  70                  75                  80

Val Cys Glu Glu Met Arg Gln Leu Phe Tyr Asp Pro Asp Glu Cys Gly
                85                  90                  95

Leu Met Lys Lys Ala Ala Leu Tyr Phe Ser Asp Phe Trp Asn Lys Leu
            100                 105                 110

Asp Val Gly Ala Ile Leu Leu Phe Val Ala Gly Leu Thr Cys Arg Leu
        115                 120                 125

Ile Pro Ala Thr Leu Tyr Pro Gly Arg Val Ile Leu Ser Leu Asp Phe
    130                 135                 140

Ile Leu Phe Cys Leu Arg Leu Met His Ile Phe Thr Ile Ser Lys Thr
145                 150                 155                 160

Leu Gly Pro Lys Ile Ile Ile Val Lys Arg Met Met Lys Asp Val Phe
                165                 170                 175

Phe Phe Leu Phe Leu Leu Ala Val Trp Val Val Ser Phe Gly Val Ala
            180                 185                 190

Lys Gln Ala Ile Leu Ile His Asn Glu Arg Arg Val Asp Trp Leu Phe
        195                 200                 205

Arg Gly Ala Val Tyr His Ser Tyr Leu Thr Ile Phe Gly Gln Ile Pro
```

210 215 220

Gly Tyr Ile Asp Gly Val Asn Phe Asn Pro Glu His Cys Ser Pro Asn
225 230 235 240

Gly Thr Asp Pro Tyr Lys Pro Lys Cys Pro Glu Ser Asp Ala Thr Gln
245 250 255

Gln Arg Pro Ala Phe Pro Glu Trp Leu Thr Val Leu Leu Leu Cys Leu
260 265 270

Tyr Leu Leu Phe Thr Asn Ile Leu Leu Leu Asn Leu Leu Ile Ala Met
275 280 285

Phe Asn Tyr Thr Phe Gln Gln Val Gln Glu His Thr Asp Gln Ile Trp
290 295 300

Lys Phe Gln Arg His Asp Leu Xaa Glu Glu Tyr His Gly Arg Pro Ala
305 310 315 320

Ala Pro Pro Pro Phe Ile Leu Leu Ser His Leu Gln Leu Phe Ile Lys
325 330 335

Arg Val Val Leu Lys Thr Pro Ala Lys Arg His Lys Gln Leu Lys Asn
340 345 350

Lys Leu Glu Lys Asn Glu Glu Ala Ala Leu Leu Ser Trp Glu Ile Tyr
355 360 365

Leu Lys Glu Asn Tyr Phe Gln Asn Arg Gln Phe Gln Gln Lys Gln Arg
370 375 380

Pro Glu Gln Lys Ile Glu Asp Ile Ser Asn Lys Val Asp Ala Met Val
385 390 395 400

Asp Leu Leu Asp Leu Asp Pro Leu Lys Arg Ser Gly Ser Met Glu Gln
405 410 415

Arg Leu Ala Ser Leu Glu Glu Gln Val Ala Gln Thr Ala Arg Ala Leu
420 425 430

His Trp Ile Val Arg Thr Leu Arg Ala Ser Gly Phe Ser Ser Glu Ala
435 440 445

Asp Val Pro Thr Leu Ala Ser Gln Lys Ala Ala Glu Glu Pro Asp Ala
450 455 460

Glu Pro Gly Gly Arg Lys Lys Thr Glu Glu Pro Gly Asp Ser Tyr His
465 470 475 480

Val Asn Ala Arg His Leu Leu Tyr Pro Asn Cys Pro Val Thr Arg Phe
485 490 495

Pro Cys Pro Thr Arg Arg Cys Pro Gly Arg Arg Ser Ser
500 505

<210> 95

<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (48)
<223> Xaa equals stop translation

<400> 95
Met Ala Gly Thr Val Leu Gly Val Gly Ala Gly Val Phe Ile Leu Ala
1 5 10 15

Leu Leu Trp Val Ala Val Leu Leu Leu Cys Val Leu Leu Ser Arg Ala
20 25 30

Ser Gly Ala Ala Arg Phe Ser Val Ile Phe Tyr Ser Ser Val Leu Xaa
35 40 45

<210> 96
<211> 293
<212> PRT
<213> Homo sapiens

<400> 96
Met Ala Thr Ala Arg Pro Pro Trp Met Trp Val Leu Cys Ala Leu Ile
1 5 10 15

Thr Ala Leu Leu Leu Gly Val Thr Glu His Val Leu Ala Asn Asn Asp
20 25 30

Val Ser Cys Asp His Pro Ser Asn Thr Val Pro Ser Gly Ser Asn Gln
35 40 45

Asp Leu Gly Ala Gly Ala Gly Glu Asp Ala Arg Ser Asp Asp Ser Ser
50 55 60

Ser Arg Ile Ile Asn Gly Ser Asp Cys Asp Met His Thr Gln Pro Trp
65 70 75 80

Gln Ala Ala Leu Leu Leu Arg Pro Asn Gln Leu Tyr Cys Gly Ala Val
85 90 95

Leu Val His Pro Gln Trp Leu Leu Thr Ala Ala His Cys Arg Lys Lys
100 105 110

Val Phe Arg Val Arg Leu Gly His Tyr Ser Leu Ser Pro Val Tyr Glu
115 120 125

Ser Gly Gln Gln Met Phe Gln Gly Val Lys Ser Ile Pro His Pro Gly
130 135 140

Tyr Ser His Pro Gly His Ser Asn Asp Leu Met Leu Ile Lys Leu Asn
145 150 155 160

```
Arg Arg Ile Arg Pro Thr Lys Asp Val Arg Pro Ile Asn Val Ser Ser
                165                 170                 175

His Cys Pro Ser Ala Gly Thr Lys Cys Leu Val Ser Gly Trp Gly Thr
            180                 185                 190

Thr Lys Ser Pro Gln Val His Phe Pro Lys Val Leu Gln Cys Leu Asn
        195                 200                 205

Ile Ser Val Leu Ser Gln Lys Arg Cys Glu Asp Ala Tyr Pro Arg Gln
    210                 215                 220

Ile Asp Asp Thr Met Phe Cys Ala Gly Asp Lys Ala Gly Arg Asp Ser
225                 230                 235                 240

Cys Gln Gly Asp Ser Gly Gly Pro Val Val Cys Asn Gly Ser Leu Gln
                245                 250                 255

Gly Leu Val Ser Trp Gly Asp Tyr Pro Cys Ala Arg Pro Asn Arg Pro
            260                 265                 270

Gly Val Tyr Thr Asn Leu Cys Lys Phe Thr Lys Trp Ile Gln Glu Thr
        275                 280                 285

Ile Gln Ala Asn Ser
    290
```

<210> 97
<211> 62
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (62)
<223> Xaa equals stop translation

<400> 97

```
Met Ala Thr Ala Arg Pro Pro Trp Met Trp Val Leu Cys Ala Leu Ile
  1               5                  10                  15

Thr Ala Leu Leu Leu Gly Val Thr Glu His Val Leu Ala Asn Asn Asp
             20                  25                  30

Val Ser Cys Asp His Pro Ser Asn Thr Val Pro Ser Gly Ser Asn Gln
         35                  40                  45

Asp Leu Gly Ala Gly Ala Gly Gly Arg Arg Pro Val Gly Xaa
     50                  55                  60
```

<210> 98
<211> 89
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (89)
<223> Xaa equals stop translation

<400> 98
Met Val Ala Gly Phe Val Phe Tyr Leu Gly Val Phe Val Val Cys His
1 5 10 15

Gln Leu Ser Ser Ser Leu Asn Ala Thr Tyr Arg Ser Leu Val Ala Arg
20 25 30

Glu Lys Val Phe Trp Asp Leu Ala Ala Thr Arg Ala Val Phe Gly Val
35 40 45

Gln Ser Thr Ala Ala Ala Val Gly Ser Ala Gly Gly Pro Cys Ala Ala
50 55 60

Cys Arg Gln Gly Ala Trp Pro Ala Glu Leu Val Leu Val Ser His His
65 70 75 80

Asp Ser Asn Gly Ile Leu Leu Leu Xaa
85

<210> 99
<211> 132
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (132)
<223> Xaa equals stop translation

<400> 99
Met Arg Gly Ser Val Glu Cys Thr Trp Gly Trp Gly His Cys Ala Pro
1 5 10 15

Ser Pro Leu Leu Leu Trp Thr Leu Leu Leu Phe Ala Ala Pro Phe Gly
20 25 30

Leu Leu Gly Glu Lys Thr Arg Gln Leu Leu Glu Phe Asp Ser Thr Asn
35 40 45

Val Ser Asp Thr Ala Ala Lys Pro Leu Gly Arg Pro Tyr Pro Pro Tyr
50 55 60

Ser Leu Ala Asp Phe Ser Trp Asn Asn Ile Thr Asp Ser Leu Asp Pro
65 70 75 80

Ala Thr Leu Ser Ala Thr Phe Gln Gly His Pro Met Asn Asp Pro Thr
85 90 95

Arg Thr Phe Ala Asn Gly Ser Leu Ala Phe Arg Ser Arg Pro Phe Pro
100 105 110

Gly Pro Ala Asp Gln Pro Asn Pro Leu Ala Ser Cys Thr Gln Gln Thr

115 120 125

Pro Val Ser Xaa
130

<210> 100
<211> 212
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (99)
<223> Xaa equals any of the naturally occurring L-amino acids

<220>
<221> SITE
<222> (212)
<223> Xaa equals stop translation

<400> 100
Met Ala Asn Ala Gly Leu Gln Leu Leu Gly Phe Ile Leu Ala Phe Leu
1 5 10 15

Gly Trp Ile Gly Ala Ile Val Ser Thr Ala Leu Pro Gln Trp Arg Ile
20 25 30

Tyr Ser Tyr Ala Gly Asp Asn Ile Val Thr Ala Gln Ala Met Tyr Glu
35 40 45

Gly Leu Trp Met Ser Cys Val Ser Gln Ser Thr Gly Gln Ile Gln Cys
50 55 60

Lys Val Phe Asp Ser Leu Leu Asn Leu Ser Ser Thr Leu Gln Ala Thr
65 70 75 80

Arg Ala Leu Met Val Val Gly Ile Leu Leu Gly Val Ile Ala Ile Phe
85 90 95

Val Ala Xaa Val Gly Met Lys Cys Met Lys Cys Leu Glu Asp Asp Glu
100 105 110

Val Gln Lys Met Arg Met Ala Val Ile Gly Gly Ala Ile Phe Leu Leu
115 120 125

Ala Gly Leu Ala Ile Leu Val Ala Thr Ala Trp Tyr Gly Asn Arg Ile
130 135 140

Val Gln Glu Phe Tyr Asp Pro Met Thr Pro Val Asn Ala Arg Tyr Glu
145 150 155 160

Phe Gly Gln Ala Leu Phe Thr Gly Trp Ala Ala Ala Ser Leu Cys Leu
165 170 175

Leu Gly Gly Ala Leu Leu Cys Cys Ser Cys Pro Arg Lys Thr Thr Ser
180 185 190

Tyr Pro Thr Pro Arg Pro Tyr Pro Lys Pro Ala Pro Ser Ser Gly Lys
195 200 205

Asp Tyr Val Xaa
210

<210> 101
<211> 9
<212> PRT
<213> Homo sapiens

<400> 101
Ile Lys Ile Ser Leu Lys Lys Arg Ser
1 5

<210> 102
<211> 79
<212> PRT
<213> Homo sapiens

<400> 102
Gly Thr Arg Gly Leu Ser Thr Val Ser Trp Thr His Thr Gln Pro Ser
1 5 10 15

Lys Arg Gly Asp Pro Ser Arg Glu Pro Arg Gly Gly His Ser Cys Leu
20 25 30

Leu Pro Gly Ser Pro Ala Thr Trp Cys Leu Pro Ala Pro Cys Ser Leu
35 40 45

Pro Gly Pro Val Leu Thr Pro Ser Ser Ser Gly Leu Asp Ser Ala Leu
50 55 60

Glu Gly Pro Arg Gly Ala Ala Ser Leu Leu Arg Ala Pro Leu Gln
65 70 75

<210> 103
<211> 23
<212> PRT
<213> Homo sapiens

<400> 103
His Thr Gln Pro Ser Lys Arg Gly Asp Pro Ser Arg Glu Pro Arg Gly
1 5 10 15

Gly His Ser Cys Leu Leu Pro
20

<210> 104
<211> 22
<212> PRT
<213> Homo sapiens

<400> 104

Val Leu Thr Pro Ser Ser Ser Gly Leu Asp Ser Ala Leu Glu Gly Pro
1 5 10 15

Arg Gly Ala Ala Ser Leu
20

<210> 105
<211> 11
<212> PRT
<213> Homo sapiens

<400> 105
Ala Gly Ser Arg Thr Asn Asn Glu Gln Ile Glu
1 5 10

<210> 106
<211> 16
<212> PRT
<213> Homo sapiens

<400> 106
Gly Thr Ser Thr Ser Ser Arg Gly Arg Leu His Ala Cys Gly His Ser
1 5 10 15

<210> 107
<211> 95
<212> PRT
<213> Homo sapiens

<400> 107
Pro Ser Ser Glu Val Gln Lys Gly Lys Pro Asn Ser Pro Leu Gly Asn
1 5 10 15

Ser Glu Leu Arg Pro His Leu Val Asn Thr Lys Pro Arg Thr Ser Leu
20 25 30

Glu Arg Gly His Thr Ile Pro Phe Leu Trp Pro Ser Glu Phe Gly Leu
35 40 45

Ser Gln Leu Trp Gly Thr Pro Ser Leu Asn Pro Asn Lys Thr Pro Leu
50 55 60

Glu Ser Leu Ser Leu His Pro Ser Pro Leu Pro Ser Ala Leu Ile Ala
65 70 75 80

Ala Arg Ile Val Thr Pro Asn Leu Thr Leu Ser Ser Leu Ile Lys
85 90 95

<210> 108
<211> 21
<212> PRT

<213> Homo sapiens

<400> 108
Pro Asn Ser Pro Leu Gly Asn Ser Glu Leu Arg Pro His Leu Val Asn
1 5 10 15

Thr Lys Pro Arg Thr
20

<210> 109
<211> 23
<212> PRT
<213> Homo sapiens

<400> 109
Leu Ser Leu His Pro Ser Pro Leu Pro Ser Ala Leu Ile Ala Ala Arg
1 5 10 15

Ile Val Thr Pro Asn Leu Thr
20

<210> 110
<211> 268
<212> PRT
<213> Homo sapiens

<400> 110
Pro Gly Ser Gln Gly Ala Ala Ala Gly Arg Glu Leu Phe Met Thr Asp
1 5 10 15

Arg Glu Arg Leu Ala Glu Ala Arg Gln Arg Glu Leu Gln Arg Gln Glu
20 25 30

Leu Leu Met Gln Lys Arg Leu Ala Met Glu Ser Asn Lys Ile Leu Gln
35 40 45

Glu Gln Gln Glu Met Glu Arg Gln Arg Arg Lys Glu Ile Ala Gln Lys
50 55 60

Ala Ala Glu Glu Asn Glu Arg Tyr Arg Lys Glu Met Glu Gln Ile Val
65 70 75 80

Glu Glu Glu Glu Lys Phe Lys Lys Gln Trp Glu Glu Asp Trp Gly Ser
85 90 95

Lys Glu Gln Leu Leu Leu Pro Lys Thr Ile Thr Ala Glu Val His Pro
100 105 110

Val Pro Leu Arg Lys Pro Lys Tyr Asp Gln Gly Val Glu Pro Glu Leu
115 120 125

Glu Pro Ala Asp Asp Leu Asp Gly Gly Thr Glu Glu Gln Gly Glu Gln
130 135 140

Asp Phe Arg Lys Tyr Glu Glu Gly Phe Asp Pro Tyr Ser Met Phe Thr
145 150 155 160

```
Pro Glu Gln Ile Met Gly Lys Asp Val Arg Leu Leu Arg Ile Lys Lys
                165                 170                 175

Glu Gly Ser Leu Asp Leu Ala Leu Glu Gly Gly Val Asp Ser Pro Ile
            180                 185                 190

Gly Lys Val Val Val Ser Ala Val Tyr Glu Arg Gly Ala Ala Glu Arg
        195                 200                 205

His Gly Gly Ile Val Lys Gly Asp Glu Ile Met Ala Ile Asn Gly Lys
    210                 215                 220

Ile Val Thr Asp Tyr Thr Leu Ala Glu Ala Asp Ala Ala Leu Gln Lys
225                 230                 235                 240

Ala Trp Asn Gln Gly Gly Asp Trp Ile Asp Leu Val Val Ala Val Cys
                245                 250                 255

Pro Pro Lys Glu Tyr Asp Asp Glu Leu Thr Phe Phe
            260                 265


<210> 111
<211> 23
<212> PRT
<213> Homo sapiens

<400> 111
Gly Arg Glu Leu Phe Met Thr Asp Arg Glu Arg Leu Ala Glu Ala Arg
  1               5                  10                  15

Gln Arg Glu Leu Gln Arg Gln
             20


<210> 112
<211> 22
<212> PRT
<213> Homo sapiens

<400> 112
Gln Gln Glu Met Glu Arg Gln Arg Arg Lys Glu Ile Ala Gln Lys Ala
  1               5                  10                  15

Ala Glu Glu Asn Glu Arg
             20


<210> 113
<211> 25
<212> PRT
<213> Homo sapiens

<400> 113
Lys Pro Lys Tyr Asp Gln Gly Val Glu Pro Glu Leu Glu Pro Ala Asp
  1               5                  10                  15
```

Asp Leu Asp Gly Gly Thr Glu Glu Gln
20 25

<210> 114
<211> 25
<212> PRT
<213> Homo sapiens

<400> 114
Ile Val Thr Asp Tyr Thr Leu Ala Glu Ala Asp Ala Ala Leu Gln Lys
1 5 10 15

Ala Trp Asn Gln Gly Gly Asp Trp Ile
20 25

<210> 115
<211> 33
<212> PRT
<213> Homo sapiens

<400> 115
His Thr Met Leu Pro Leu Lys Ile Ala Ala Pro Tyr Leu Leu Glu Asn
1 5 10 15

Cys Ser Cys Pro Ile Tyr Ile Ser Thr Ser Pro His Leu Phe Leu Ser
20 25 30

Thr

<210> 116
<211> 19
<212> PRT
<213> Homo sapiens

<400> 116
Phe Ser Ile Leu Phe Ala Phe Val Leu Phe Tyr Pro Gly Ser Phe Phe
1 5 10 15

Thr Leu Pro

<210> 117
<211> 6
<212> PRT
<213> Homo sapiens

<400> 117
His Glu Ser Thr Val Lys
1 5

<210> 118
<211> 27

<212> PRT
<213> Homo sapiens

<400> 118
Leu Glu Asn Leu Gly Thr His Lys Lys Lys Asp Ser Phe Ser Val Lys
1 5 10 15

Thr Val Gly Ile Cys Cys Cys Phe His Leu Asn
20 25

<210> 119
<211> 6
<212> PRT
<213> Homo sapiens

<400> 119
Phe Thr Lys Cys Phe His
1 5

<210> 120
<211> 8
<212> PRT
<213> Homo sapiens

<400> 120
Gln Asn Met Asn Asp Tyr Asn Ile
1 5

<210> 121
<211> 51
<212> PRT
<213> Homo sapiens

<400> 121
Pro Ala Arg His Leu Trp Thr Pro Ser Pro Val Cys Lys Pro Ser Ile
1 5 10 15

Lys Pro His Val Ser Phe Ala Gly Ser Gly Ser Leu Trp Arg Leu Glu
20 25 30

Pro Tyr Ala Phe Pro Ile Glu Val Asn Arg Gly Ser Ala Gln His Trp
35 40 45

Val Pro Gly
50

<210> 122
<211> 29
<212> PRT
<213> Homo sapiens

<400> 122
Val Cys Lys Pro Ser Ile Lys Pro His Val Ser Phe Ala Gly Ser Gly
1 5 10 15

Ser Leu Trp Arg Leu Glu Pro Tyr Ala Phe Pro Ile Glu
20 25

<210> 123
<211> 32
<212> PRT
<213> Homo sapiens

<400> 123
Gln Phe Ser Phe Leu Ser Ala Lys Gly Leu His Trp Ala Leu Phe Val
1 5 10 15

Phe Phe Tyr Phe Leu Ser Thr Ala Cys Gln Arg Trp Ala Trp Gly Leu
20 25 30

<210> 124
<211> 82
<212> PRT
<213> Homo sapiens

<400> 124
His Glu Pro Gly Arg Cys Gly Pro Glu Asn Leu Ala Leu Gln Ala Thr
1 5 10 15

Gln Arg Gly Thr Arg Phe Ser Val Pro Met Cys Lys Ser Ser Arg Gln
20 25 30

Tyr Thr Tyr Thr Ser Val His Met Cys Gln Cys Ala Cys Glu Arg Val
35 40 45

Glu Trp Arg Gly Ser Leu Thr Pro Ala Arg Ala Leu His Asn His Leu
50 55 60

Thr Glu Gln Trp Phe Pro His Gly Phe Pro Phe Leu Ser Arg Phe Phe
65 70 75 80

Thr Tyr

<210> 125
<211> 24
<212> PRT
<213> Homo sapiens

<400> 125
Glu Asn Leu Ala Leu Gln Ala Thr Gln Arg Gly Thr Arg Phe Ser Val
1 5 10 15

Pro Met Cys Lys Ser Ser Arg Gln
20

<210> 126
<211> 26
<212> PRT
<213> Homo sapiens

<400> 126
Met Cys Gln Cys Ala Cys Glu Arg Val Glu Trp Arg Gly Ser Leu Thr
1 5 10 15

Pro Ala Arg Ala Leu His Asn His Leu Thr
20 25

<210> 127
<211> 12
<212> PRT
<213> Homo sapiens

<400> 127
Leu Arg Arg Ala Ser Cys Pro Ile Trp Ser Lys Asp
1 5 10

<210> 128
<211> 166
<212> PRT
<213> Homo sapiens

<400> 128
Gly Thr Ser Thr Lys Leu Pro Tyr Cys Arg Glu Asn Val Cys Leu Ala
1 5 10 15

Tyr Gly Ser Glu Trp Ser Val Tyr Ala Val Gly Ser Gln Ala His Val
20 25 30

Ser Phe Leu Asp Pro Arg Gln Pro Ser Tyr Asn Val Lys Ser Val Cys
35 40 45

Ser Arg Glu Arg Gly Ser Gly Ile Arg Ser Val Ser Phe Tyr Glu His
50 55 60

Ile Ile Thr Val Gly Thr Gly Gln Gly Ser Leu Leu Phe Tyr Asp Ile
65 70 75 80

Arg Ala Gln Arg Phe Leu Glu Glu Arg Leu Ser Ala Cys Tyr Gly Ser
85 90 95

Lys Pro Arg Leu Ala Gly Glu Asn Leu Lys Leu Thr Thr Gly Lys Gly
100 105 110

Trp Leu Asn His Asp Glu Thr Trp Arg Asn Tyr Phe Ser Asp Ile Asp
115 120 125

Phe Phe Pro Asn Ala Val Tyr Thr His Cys Tyr Asp Ser Ser Gly Thr
130 135 140

Lys Leu Phe Val Ala Gly Gly Pro Leu Pro Ser Gly Leu His Gly Asn

145 150 155 160

Tyr Ala Gly Leu Trp Ser
165

<210> 129
<211> 22
<212> PRT
<213> Homo sapiens

<400> 129
Cys Arg Glu Asn Val Cys Leu Ala Tyr Gly Ser Glu Trp Ser Val Tyr
1 5 10 15

Ala Val Gly Ser Gln Ala
20

<210> 130
<211> 24
<212> PRT
<213> Homo sapiens

<400> 130
Pro Ser Tyr Asn Val Lys Ser Val Cys Ser Arg Glu Arg Gly Ser Gly
1 5 10 15

Ile Arg Ser Val Ser Phe Tyr Glu
20

<210> 131
<211> 29
<212> PRT
<213> Homo sapiens

<400> 131
Asp Ile Arg Ala Gln Arg Phe Leu Glu Glu Arg Leu Ser Ala Cys Tyr
1 5 10 15

Gly Ser Lys Pro Arg Leu Ala Gly Glu Asn Leu Lys Leu
20 25

<210> 132
<211> 26
<212> PRT
<213> Homo sapiens

<400> 132
Lys Leu Thr Thr Gly Lys Gly Trp Leu Asn His Asp Glu Thr Trp Arg
1 5 10 15

Asn Tyr Phe Ser Asp Ile Asp Phe Phe Pro
20 25

<210> 133
<211> 21
<212> PRT
<213> Homo sapiens

<400> 133
Tyr Asp Ser Ser Gly Thr Lys Leu Phe Val Ala Gly Gly Pro Leu Pro
1 5 10 15

Ser Gly Leu His Gly
20

<210> 134
<211> 280
<212> PRT
<213> Homo sapiens

<400> 134
Lys Pro Gln Arg Phe Arg Arg Pro Phe Phe Phe Asn His Pro Lys Pro
1 5 10 15

Ser Ser His Pro Gly Leu His Ser Arg Pro Thr Leu His Ser His Pro
20 25 30

Ala Phe His Ser His Pro Glu Leu Gln Gln Pro Thr Gln Thr Ser Pro
35 40 45

Val Pro Leu Thr Pro Glu Ser Pro Leu Phe Gln Asn Phe Ser Gly Tyr
50 55 60

His Ile Gly Val Gly Arg Ala Asp Cys Thr Gly Gln Val Ala Asp Ile
65 70 75 80

Asn Leu Met Gly Tyr Gly Lys Ser Gly Gln Asn Ala Gln Gly Ile Leu
85 90 95

Thr Arg Leu Tyr Ser Arg Ala Phe Ile Met Ala Glu Pro Asp Gly Ser
100 105 110

Asn Arg Thr Val Phe Val Ser Ile Asp Ile Gly Met Val Ser Gln Arg
115 120 125

Leu Arg Leu Glu Val Leu Asn Arg Leu Gln Ser Lys Tyr Gly Ser Leu
130 135 140

Tyr Arg Arg Asp Asn Val Ile Leu Ser Gly Thr His Thr His Ser Gly
145 150 155 160

Pro Ala Gly Tyr Phe Gln Tyr Thr Val Phe Val Ile Ala Ser Glu Gly
165 170 175

Phe Ser Asn Gln Thr Phe Gln His Met Val Thr Gly Ile Leu Lys Ser
180 185 190

Ile Asp Ile Ala His Thr Asn Met Lys Pro Gly Lys Ile Phe Ile Asn
195 200 205

```
Lys Gly Asn Val Asp Gly Val Gln Ile Asn Arg Ser Pro Tyr Ser Tyr
    210                 215                 220

Leu Gln Asn Pro Gln Ser Glu Arg Ala Arg Tyr Ser Ser Asn Thr Asp
225                 230                 235                 240

Lys Glu Met Ile Val Leu Lys Met Val Asp Leu Asn Gly Asp Asp Leu
                245                 250                 255

Gly Leu Ile Ser Phe Ser Phe Ser Lys Ser Ala Leu Gly Thr Tyr Tyr
            260                 265                 270

Glu Pro Arg Asn Thr Ser Leu Glu
        275                 280


<210> 135
<211> 30
<212> PRT
<213> Homo sapiens

<400> 135
Lys Pro Ser Ser His Pro Gly Leu His Ser Arg Pro Thr Leu His Ser
  1               5                  10                  15

His Pro Ala Phe His Ser His Pro Glu Leu Gln Gln Pro Thr
             20                  25                  30


<210> 136
<211> 26
<212> PRT
<213> Homo sapiens

<400> 136
Arg Ala Asp Cys Thr Gly Gln Val Ala Asp Ile Asn Leu Met Gly Tyr
  1               5                  10                  15

Gly Lys Ser Gly Gln Asn Ala Gln Gly Ile
             20                  25


<210> 137
<211> 24
<212> PRT
<213> Homo sapiens

<400> 137
Arg Ala Phe Ile Met Ala Glu Pro Asp Gly Ser Asn Arg Thr Val Phe
  1               5                  10                  15

Val Ser Ile Asp Ile Gly Met Val
             20


<210> 138
<211> 27
<212> PRT
```

<213> Homo sapiens

<400> 138
Arg Leu Gln Ser Lys Tyr Gly Ser Leu Tyr Arg Arg Asp Asn Val Ile
1 5 10 15

Leu Ser Gly Thr His Thr His Ser Gly Pro Ala
20 25

<210> 139
<211> 23
<212> PRT
<213> Homo sapiens

<400> 139
Ala Ser Glu Gly Phe Ser Asn Gln Thr Phe Gln His Met Val Thr Gly
1 5 10 15

Ile Leu Lys Ser Ile Asp Ile
20

<210> 140
<211> 24
<212> PRT
<213> Homo sapiens

<400> 140
Ile Phe Ile Asn Lys Gly Asn Val Asp Gly Val Gln Ile Asn Arg Ser
1 5 10 15

Pro Tyr Ser Tyr Leu Gln Asn Pro
20

<210> 141
<211> 30
<212> PRT
<213> Homo sapiens

<400> 141
Thr Asp Lys Glu Met Ile Val Leu Lys Met Val Asp Leu Asn Gly Asp
1 5 10 15

Asp Leu Gly Leu Ile Ser Phe Ser Phe Ser Lys Ser Ala Leu
20 25 30

<210> 142
<211> 15
<212> PRT
<213> Homo sapiens

<400> 142
His Leu Thr Gln Phe Cys Val Ile Gln Leu Leu Pro Thr His Leu
1 5 10 15

<210> 143
<211> 105
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (5)
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 143
Met Leu Ala Ser Xaa Ser Val Leu Cys Asp Pro Ala Pro Ala Asn Pro
1 5 10 15

Ser Asp Glu Leu Leu Val His Ser Pro Cys Phe Leu Leu Ser Pro Pro
20 25 30

Val Ala Pro Val Pro Phe Phe Pro Cys Ala Lys Leu Ile Pro Ala Pro
35 40 45

Arg Pro Val Arg Tyr Phe Ser Pro Pro Asp Leu Arg Leu Gly Asn Thr
50 55 60

Pro Ala Pro Ser Glu Ile Thr Tyr Thr Pro Ser Val His Phe Cys His
65 70 75 80

Pro Ile Ala Lys Leu Leu Cys Leu Lys Val Arg Asn Leu Cys Glu Gly
85 90 95

Val Leu Ser Ala Ala Phe Pro Lys Ala
100 105

<210> 144
<211> 27
<212> PRT
<213> Homo sapiens

<400> 144
Ala Asn Pro Ser Asp Glu Leu Leu Val His Ser Pro Cys Phe Leu Leu
1 5 10 15

Ser Pro Pro Val Ala Pro Val Pro Phe Phe Pro
20 25

<210> 145
<211> 34
<212> PRT
<213> Homo sapiens

<400> 145
Phe Ser Pro Pro Asp Leu Arg Leu Gly Asn Thr Pro Ala Pro Ser Glu
1 5 10 15

Ile Thr Tyr Thr Pro Ser Val His Phe Cys His Pro Ile Ala Lys Leu
20 25 30

Leu Cys

<210> 146
<211> 8
<212> PRT
<213> Homo sapiens

<400> 146
Leu Ile Phe His Phe Val Tyr Leu
1 5

<210> 147
<211> 173
<212> PRT
<213> Homo sapiens

<400> 147
Pro Thr Pro Arg Val Ile Leu Gln Val Gly Ser Arg Ile Ala Asp Arg
1 5 10 15

Val Tyr Asp Ile Pro Arg Asn Phe Pro Leu Ala Leu Asp Leu Gly Cys
20 25 30

Gly Arg Gly Tyr Ile Ala Gln Tyr Leu Asn Lys Leu Gln Leu Phe His
35 40 45

Cys Arg Lys Leu Leu Glu Ser Phe Ser Lys Leu Thr Leu Gln Lys Met
50 55 60

Leu Cys Leu His Trp Val Asn Asp Leu Pro Arg Ala Leu Glu Gln Ile
65 70 75 80

His Tyr Ile Leu Lys Pro Asp Gly Val Phe Ile Gly Ala Met Phe Gly
85 90 95

Gly Asp Thr Leu Tyr Glu Leu Arg Cys Ser Leu Gln Leu Ala Glu Thr
100 105 110

Glu Arg Glu Gly Gly Phe Ser Pro His Ile Ser Pro Phe Thr Ala Val
115 120 125

Asn Asp Leu Gly His Leu Leu Gly Arg Ala Gly Phe Asn Thr Leu Thr
130 135 140

Val Asp Thr Asp Glu Ile Gln Val Asn Tyr Pro Gly Met Phe Glu Leu
145 150 155 160

Met Glu Asp Leu Gln Glu Gln Lys Ser Arg Met Leu Thr
165 170

<210> 148
<211> 24
<212> PRT

<213> Homo sapiens

<400> 148
Leu Gln Val Gly Ser Arg Ile Ala Asp Arg Val Tyr Asp Ile Pro Arg
1 5 10 15

Asn Phe Pro Leu Ala Leu Asp Leu
20

<210> 149
<211> 24
<212> PRT
<213> Homo sapiens

<400> 149
Gly Tyr Ile Ala Gln Tyr Leu Asn Lys Leu Gln Leu Phe His Cys Arg
1 5 10 15

Lys Leu Leu Glu Ser Phe Ser Lys
20

<210> 150
<211> 27
<212> PRT
<213> Homo sapiens

<400> 150
Val Asn Asp Leu Pro Arg Ala Leu Glu Gln Ile His Tyr Ile Leu Lys
1 5 10 15

Pro Asp Gly Val Phe Ile Gly Ala Met Phe Gly
20 25

<210> 151
<211> 28
<212> PRT
<213> Homo sapiens

<400> 151
Tyr Glu Leu Arg Cys Ser Leu Gln Leu Ala Glu Thr Glu Arg Glu Gly
1 5 10 15

Gly Phe Ser Pro His Ile Ser Pro Phe Thr Ala Val
20 25

<210> 152
<211> 22
<212> PRT
<213> Homo sapiens

<400> 152
Asn Thr Leu Thr Val Asp Thr Asp Glu Ile Gln Val Asn Tyr Pro Gly
1 5 10 15

Met Phe Glu Leu Met Glu
20

<210> 153
<211> 86
<212> PRT
<213> Homo sapiens

<400> 153
Met Arg Gln Leu Phe Tyr Asp Pro Asp Glu Cys Gly Leu Met Lys Lys
1 5 10 15

Gly Gly Leu Tyr Phe Ser Asp Phe Trp Asn Lys Leu Asp Val Gly Ala
20 25 30

Ile Leu Leu Phe Val Ala Gly Leu Thr Cys Arg Leu Ile Pro Ala Thr
35 40 45

Leu Tyr Pro Gly Arg Val Ile Leu Ser Leu Asp Phe Ile Leu Phe Cys
50 55 60

Leu Arg Leu Met His Ile Phe Thr Ile Ser Lys Thr Leu Gly Pro Lys
65 70 75 80

Ile Ile Ile Val Lys Arg
85

<210> 154
<211> 27
<212> PRT
<213> Homo sapiens

<400> 154
Asp Glu Cys Gly Leu Met Lys Lys Gly Gly Leu Tyr Phe Ser Asp Phe
1 5 10 15

Trp Asn Lys Leu Asp Val Gly Ala Ile Leu Leu
20 25

<210> 155
<211> 25
<212> PRT
<213> Homo sapiens

<400> 155
Thr Leu Tyr Pro Gly Arg Val Ile Leu Ser Leu Asp Phe Ile Leu Phe
1 5 10 15

Cys Leu Arg Leu Met His Ile Phe Thr
20 25

<210> 156
<211> 274
<212> PRT

<213> Homo sapiens

<400> 156

```
Val Pro Arg Glu Arg Arg Asp Ala Ala Glu Pro Ala Phe Pro Glu Trp
  1               5                  10                  15

Leu Thr Val Leu Leu Leu Cys Leu Tyr Leu Leu Phe Thr Asn Ile Leu
             20                  25                  30

Leu Leu Asn Leu Leu Ile Ala Met Phe Asn Tyr Thr Phe Gln Gln Val
         35                  40                  45

Gln Glu His Thr Asp Gln Ile Trp Lys Phe Gln Arg His Asp Leu Ile
     50                  55                  60

Glu Glu Tyr His Gly Arg Pro Ala Val Pro Pro Pro Leu Ile Leu Phe
 65                  70                  75                  80

Ser His Leu Gln Leu Phe Ile Lys Arg Val Val Leu Lys Thr Pro Ala
                 85                  90                  95

Lys Arg His Lys Gln Leu Lys Asn Lys Leu Glu Lys Asn Glu Glu Ala
            100                 105                 110

Ala Leu Leu Ser Trp Glu Ile Tyr Leu Lys Glu Asn Tyr Leu Gln Asn
        115                 120                 125

Arg Gln Phe Gln Gln Lys Gln Arg Pro Glu Gln Lys Ile Glu Asp Ile
    130                 135                 140

Ser Asn Lys Val Asp Ala Met Val Asp Leu Leu Asp Leu Asp Pro Leu
145                 150                 155                 160

Lys Arg Ser Gly Ser Met Glu Gln Arg Leu Ala Ser Leu Glu Glu Gln
                165                 170                 175

Val Ala Gln Thr Ala Arg Ala Leu His Trp Ile Val Arg Thr Leu Arg
            180                 185                 190

Ala Ser Gly Phe Ser Ser Glu Ala Asp Val Pro Thr Leu Ala Ser Gln
        195                 200                 205

Lys Ala Ala Glu Glu Pro Asp Ala Glu Pro Gly Gly Arg Lys Lys Thr
    210                 215                 220

Glu Glu Pro Gly Asp Ser Tyr His Val Asn Ala Arg His Leu Leu Tyr
225                 230                 235                 240

Pro Asn Cys Pro Val Thr Arg Phe Pro Val Pro Asn Glu Lys Val Pro
                245                 250                 255

Trp Glu Thr Glu Phe Leu Ile Tyr Asp Pro Pro Phe Tyr Thr Ala Glu
            260                 265                 270

Arg Lys
```

<210> 157
<211> 28
<212> PRT
<213> Homo sapiens

<400> 157
Gln Ile Trp Lys Phe Gln Arg His Asp Leu Ile Glu Glu Tyr His Gly
1 5 10 15

Arg Pro Ala Val Pro Pro Pro Leu Ile Leu Phe Ser
20 25

<210> 158
<211> 28
<212> PRT
<213> Homo sapiens

<400> 158
Leu Gln Asn Arg Gln Phe Gln Gln Lys Gln Arg Pro Glu Gln Lys Ile
1 5 10 15

Glu Asp Ile Ser Asn Lys Val Asp Ala Met Val Asp
20 25

<210> 159
<211> 24
<212> PRT
<213> Homo sapiens

<400> 159
Val Pro Thr Leu Ala Ser Gln Lys Ala Ala Glu Glu Pro Asp Ala Glu
1 5 10 15

Pro Gly Gly Arg Lys Lys Thr Glu
20

<210> 160
<211> 20
<212> PRT
<213> Homo sapiens

<400> 160
Pro Asn Glu Lys Val Pro Trp Glu Thr Glu Phe Leu Ile Tyr Asp Pro
1 5 10 15

Pro Phe Tyr Thr
20

<210> 161
<211> 48
<212> PRT
<213> Homo sapiens

<400> 161

Pro Glu Ser Phe Asn Phe Cys Phe Gly Pro Gly Val Pro Met Pro Trp
1 5 10 15

Cys Leu Leu Pro Val Leu Ser Val Leu His Trp Ser Thr Glu Asp Thr
20 25 30

Arg Ser Cys Gly Ala Gln Gly Gly Gly Pro Pro Leu Pro Pro Arg Gly
35 40 45

<210> 162
<211> 30
<212> PRT
<213> Homo sapiens

<400> 162
Leu Trp Thr Met Asn Pro Ala Ser Asp Gly Gly Thr Ser Glu Ser Ile
1 5 10 15

Phe Asp Leu Asp Tyr Ala Ser Trp Gly Ile Arg Ser Thr Leu
20 25 30

<210> 163
<211> 23
<212> PRT
<213> Homo sapiens

<400> 163
Gly Thr Ser Gly Gly Ala Arg Ala Ser Leu Gly Pro Ser Pro His Leu
1 5 10 15

His Gln Gly Pro Gly Ala Thr
20

<210> 164
<211> 9
<212> PRT
<213> Homo sapiens

<400> 164
Tyr Leu Val Ile Phe Phe Leu Lys Cys
1 5

<210> 165
<211> 7
<212> PRT
<213> Homo sapiens

<400> 165
Gly Ser Gln Val Asn Gly Val
1 5

<210> 166
<211> 10
<212> PRT
<213> Homo sapiens

<400> 166
Arg Pro Thr Arg Pro Leu Asn Cys Gly Arg
1 5 10

<210> 167
<211> 39
<212> PRT
<213> Homo sapiens

<400> 167
Val Trp Gly Pro Pro Ser Val Ala Ala Ala Leu Glu Leu Val Asp Pro
1 5 10 15

Pro Gly Cys Arg Glu Phe Gly Thr Ser Asn Ile Glu Asp Arg Asp Glu
20 25 30

Leu Ala Tyr His Ile Ser Ile
35

<210> 168
<211> 36
<212> PRT
<213> Homo sapiens

<400> 168
Leu Pro Pro Arg Gly Pro Ala Thr Phe Gly Ser Pro Gly Cys Pro Pro
1 5 10 15

Ala Asn Ser Pro Pro Ser Ala Pro Ala Thr Pro Glu Pro Ala Arg Ala
20 25 30

Pro Glu Arg Val
35

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00108

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/00; C12N 1/15, 1/21, 5/10, 15/12, 15/63

US CL :435/252.3, 254.11, 320.1, 325, 440; 536/23.1, 23.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/252.3, 254.11, 320.1, 325, 440; 536/23.1, 23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

GENBANK, EMBL, SWISS-PROT, GENESEARCH, PIR, SPTREMBL
search terms: *SEQ ID NO: 11-20 and 56-65*

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. D31303, SUDO et al. '2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library', complete record, 08 February 1995.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA482858, NCI-CGAP 'National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', complete record, 19 August 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA535611, NCI-CGAP 'National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', complete record, 21 August 1997.	1, 7-10

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
13 APRIL 1999	11 MAY 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer SCOTT D. PRIEBE Telephone No. (703) 308-0196

JOYCE BRIDGERS
PARALEGAL SPECIALIST
CHEMICAL MATRIX

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00108

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA515902, NCI-CGAP 'National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', complete record, 19 August 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA121174, HILLIER et al. 'WashU-Merck EST Project', complete record, 19 May 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA613157, NCI-CGAP 'National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', complete record, 16 October 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. B46271, MAHAIRAS et al. 'Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 sequence Tagged Connectors', complete record, 20 October 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. Z93016, BURTON, J. 'Human DNA sequence from PAC 211D12 on chromosome 20q12-13.2. Contains Krs-2, K= channel protein, stress responsive', partial record, 19 December 1997.	1-10, 21
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. AA369250, ADAMS et al. 'Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence', complete record, 21 April 1997.	1, 7-10
X	Database GenBank, US National Library of Medicine, (Bethesda, MD, USA), No. T81709, HILLIER et al. 'yd94a07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE: 115860 5', mRNA sequence', complete record, 15 March 1995.	1, 7-10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00108

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-10, 21

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00108

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
~~This ISA found multiple inventions as follows:~~

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Groups I-XVIII, claim(s) 1-10 and 21, drawn to a polynucleotide, vector comprising same, first claimed method of use, i.e. using polynucleotide to make a cell, and the cell made by the process. Claims 1-10 and 21 recite 39 independent polynucleotides (SEQ ID NO: 11-46, 51, 52 and 54 or encoding SEQ ID NO: 56-91, 96, 97 and 99). Group I consists of the first ten polynucleotides (SEQ ID NOs 11-20 or encoding SEQ ID NOs 56-65). Each of groups II-XVIII consists of up to four of the remaining 29 polynucleotides, in order.

Groups XIX-LVII, claim(s) 11, 12, 14-16 and 17 (first part), drawn to a polypeptide, a method of making the polypeptide and first claimed method of use, i.e. in treatment. These claims recite 39 independent polypeptides, each of groups XIX-LVII consists of a single polypeptide as set forth in SEQ ID NOs 56-91, 96, 97 and 99, respectively.

Groups LVIII-XCVI, claim(s) 13 and 19, drawn to an antibody to a polypeptide and the first claimed method of using same. These claims recite 39 independent antibodies to 39 independent polypeptides, each of groups LVIII-XCVI consists an antibody against a single polypeptide as set forth in SEQ ID NOs 56-91, 96, 97 and 99, respectively.

Groups XCVII-CXIV, claim(s) 17(second part), drawn to an additional method of using a polynucleotide. Group XCVII consists of methods reciting the first ten polynucleotides (SEQ ID NOs 11-20 or encoding SEQ ID NOs 56-65). Each of groups XCVIII-CXIV pertains to up to four of the remaining 29 polynucleotides, in order.

Groups CXV-CXXXII, claim(s) 18, drawn to a second additional method of using a polynucleotide. Group CXV consists of methods reciting the first ten polynucleotides (SEQ ID NOs 11-20 or encoding SEQ ID NOs 56-65). Each of groups CXVI-CXXXII pertains to up to four of the remaining 29 polynucleotides, in order.

Groups CXXXIII-CLXXI, claim(s) 20, drawn to an additional method of using the polypeptide. These claims recite 39 independent methods of using 39 independent polypeptides, each of groups CXXXIII-CLXXI consists an antibody against a single polypeptide as set forth in SEQ ID NOs 56-91, 96, 97 and 99, respectively.

Groups CLXXII-CLXXXIX, claim 22, drawn to a third additional method of using a polynucleotide. Group CLXXII consists of methods reciting the first ten polynucleotides (SEQ ID NOs 11-20 or encoding SEQ ID NOs 56-65). Each of groups CLXXIII-CLXXXIX pertains to up to four of the remaining 29 polynucleotides, in order.

Claim 23 is unsearchable and cannot be grouped as it is drawn to unknown and unspecified compounds.

The inventions listed as Groups I-CLXXXIX do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Each of the corresponding polynucleotides, polypeptides and antibodies are independent products, with different uses and being structurally, biochemically and biologically different products. In additional or alternate methods of use are claimed for individual polynucleotides and polypeptides. 37 CFR 1.475(b) does not provide for unity of invention of more than 1 product or more than one method of using a product as a combination of invention having unity of invention. However, with respect to groups drawn to independent polynucleotides or alternate methods of using same recited in the alternative, in accordance with 1192 O.G. 68 (19 November 1966) applicant is entitled to an initial search of inventions pertaining to the first ten independent polynucleotides recited, and may elect to pay an additional fee for each search of up to four additional independent polynucleotides. For additional method of using each of the independent polynucleotides, applicant may further elect to pay an additional fee for an additional search involving the first ten polynucleotides and each additional search involving up to four additional polynucleotides. With respect to groups pertaining to independent polypeptides or antibodies to the independent polypeptides, each product or method of use is an additional invention. An additional fee must be paid for search of each additional invention relating to polypeptides or antibodies against same. With respect to the relationship between the claimed polynucleotides and the claimed polypeptides, there is no one-to-one correspondence, i.e. no corresponding scope, between claims drawn to polynucleotides and their use and those drawn to polypeptides, antibodies and their use. Consequently, there is no special technical feature linking the polynucleotides and the polypeptides or antibodies claimed.

Only Group I was searched. Claims 1-10 and 21 were searched only as they pertain to SEQ ID NOs 11-20 and 56-65.